REGULATION OF CpG ISLAND PROMOTERS BY THE HISTONE METHYLTRANSFERASE MLL2

By

VASILEIOS LADOPOULOS

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Biomedical Research

School of Cancer Sciences

College of Medical and Dental Sciences

University of Birmingham

2012

Abstract

MLL2 is a H3K4 specific HMT which vital for normal embryonic development in the mouse. Little is known on how MLL2 is recruited to its target genes and activates transcription. To gain insight into the molecular mechanism underlying MLL2 function, we focused on a known MLL2 target gene: *Magoh2*. This gene is controlled by a CpG island promoter and is ubiquitously expressed.

Our results demonstrate that in the absence of MLL2, the *Magoh2* promoter is methylated and *Magoh2* is transcriptionally silenced. The *Magoh2* promoter adopts the active conformation only in the presence of MLL2. Pol II is lost from the *Magoh2* promoter in the absence of MLL2, resulting in *Magoh2* down-regulation. We observed loss of H3K4me$_3$ and H3K9ac and relocation of a nucleosome over the promoter, coinciding with the onset of DNA methylation. Use of DRB and α-amanitin demonstrated that neither transcription nor the presence of Pol II are required for the maintenance of H3K4me$_3$. *Magoh2* silencing can be overcome by re-introducing full-length MLL2.

We investigated the role of MLL2 in haemopoiesis and demonstrated that MLL2 is vital for macrophage differentiation from embryoid bodies. MLL2 may be required for correct upregulation of *Flk1* and generation of haemangioblast cells. When *Mll2* was deleted in haemangioblasts, the haemopoietic transcriptional program was perturbed suggesting that MLL2 may also play a role at this later developmental stage.

Acknowledgements

I would like to thank my supervisors Prof. Constanze Bonifer, Dr. Maarten Hoogenkamp and Prof. Peter Cockerill for giving me the opportunity to pursue a PhD degree as a member of the Bonifer lab. Their support and advice has been of paramount significance for the completion of this work. I owe most of the technical expertise I obtained during the course of my PhD studies to Dr. Maarten Hoogenkamp to whom I am very grateful.

I would like to thank all members of the Bonifer/Cockerill labs – past and present – for the intellectual stimulation and constructive criticism they provided throughout the past 4 years. In particular, I would like to thank Dr. Monika Lichtinger and Dr. Jane Gilmour for their critical review of this manuscript.

I am thankful to Prof. F. Stewart, Prof. Konstantinos Anastassiadis and all the members of their research groups for welcoming me in their laboratory in Dresden. The contributions of Mrs. Madeleine Walker and Dr. Jun Fu in the recombineering exercises described herein have been invaluable. Additionally, I would like to thank Prof. F. Stewart and Prof. H. Stunnenberg for sharing unpublished data with me.

I would like to thank Dr. Lylia Ouboussad for her support and tolerance during the past 4 stressful months and for proofreading this document. Additionally I would like to thank Dr. Giacomo Volpe and Dr. Pierre Cauchy for being much than colleaguesand making my time in Birmingham thoroughly enjoyable.

My sincerest thanks to the EuTRACC and Prof. Arthur Riggs for financially supporting this project; this work would have been impossible without their contributions.

Lastly, I would like to dedicate this to my parents and my sister for their support in this endeavour and to As. Prof. Konstantinos Vlachonassios for encouraging me to pursue an academic career in the first place.

TABLE OF CONTENTS

List of figures

List of tables

Abbreviations

AEBSF	4-(2-aminoethyl) benzenesulphonyl fluoride hydrochloride
APC	allophycocyanin
APS	ammonium persulphate
ATP	adenosine triphosphate
B&W	bind & wash
BAC	bacterial artificial chromosome
Bcl2	B-cell lymphoma 2
Bdnf	brain-derived neurotrophic factor
BLAST	basic local alignment search tool
bp	base pair
BRD4	bromodomain-containing protein 4
BSA	bovine serum albumin
CBP	CREB binding protein
Cdk9	cyclin dependent kinase 9
Cfp1	CxxC finger protein 1
ChIP	chromatin immunoprecipitation
CI	chloroform/isoamyl-alcohol
cpm	counts per minute
CREB	cAMP response element-binding
Csf1R	colony-stimulating factor 1 receptor
CTCF	CCCTC-binding factor
CTD	C-terminal domain

Cy	cyanine
dCTP	deoxycytidine triphosphate
DHS	DNaseI hypersensitive site
DMEM	Dulbecco's modified Eagle's medium
DMS	dimethyl-sulphate
DMSO	dimethyl-sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
dNTP	deoxyribonucleotide triphosphate
DPE	downstream promoter element
DRB	5, 6-dichlorobenzimidazole 1-β-D-ribofuranoside
ds	double stranded
DSIF	DRB sensitivity inducing factor
DTT	dithiothreitol
EB	embryoid bodies
EDTA	ethylene diamine tetra-acetic acid
EGTA	ethylene glycol tetra-acetic acid
EMSA	electrophoretic mobility shift assay
ES	embryonic stem
EZH2	enhancer of zeste homologue 2
FACS	fluorescence activated cell sorting
FBS	foetal bovine serum
FCS	foetal calf serum

FITC	fluorescein isothiocyanate
Flk	foetal liver kinase
FosL1	Fos-related antigen 1
G418	geneticin
GATA2	GATA binding protein 2
GCN5	general control of amino acid synthesis 5
Gfi1	growth factor independent 1
GFP	green fluorescent protein
HAT	histone acetyltransferase
HCF-1	host cell factor 1
HDAC	histone deacetylase
HE	haemogenic endothelium
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HIF	hypoxia induced factor 1
HOX	homeobox
HP1	heterochromatin protein 1
HRP	horseradish peroxidise
Hsp	heat shock protein
ICM	inner cell mass
Ig	immunoglobulin
IL	interleukin
IMDM	Iscove's modified Dulbecco's medium
INO80	inositol requiring 80
Inr	initiator

IP	immunoprecipitation
IRES	internal ribosome entry site
ISWI	imitation SWI
IVD	*in vitro* differentiation
Kb	kilo base pair
LB	Luria-Bertani
LIF	leukaemia inhibitory factor
LM-PCR	ligation mediated PCR
LSD1	lysine-specific demethylase 1
MACS	magnetism-activated cell sorting
MBD	methyl-binding domain
M-CSF	macrophage colony-stimulating factor
MEIS1	myeloid ecotropic viral integration site 1
miRNA	micro RNA
MLL	myeloid/lymphoid or mixed-lineage leukaemia
M-MLV	Moloney murine leukemia virus
MNase	Micrococcal nuclease
mRNA	messenger RNA
MTG	mono-thioglycerol
NCBI	national centre for biotechnology information
ncRNA	non-coding RNA
NELF	negative elongation factor
NP40	non-ident P40
Nup188	nucleoporin 188

OHT	4-hydroxytamoxifen
P	phenol
Paf	polymerase associated factor
PBS	phosphate buffered saline
PcG	polycomb group
PCI	phenol/chloroform/isoamyl-alcohol
PCR	polymerase chain reaction
PE	phycoerythrin
PEG	polyethylene glycol
PHD	plant homeo-domain
PhoRC	pleiohomeotic repressive complex
PI	propidium iodide
PIC	protease inhibitor cocktail
Pigp	phosphatidylinositol glycan, class P
PIM	proto-oncogene serine/threonine-protein kinase
PMSF	phenylmethylsulphonyl fluoride
PNK	polynucleotide kinase
PRC	polycomb repressive complex
p-TEFb	positive transcription elongation factor b
PTM	post-translational modification
QPCR	quantitative PCR
Rbp1	retinol binding protein 1
RE	restriction enzyme
RIPA	radio immunoprecipitation assay

RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcription
RUNX1	Runt-related transcription factor 1
sc	subclone
sd	standard deviation
SDS	sodium dodecyl-sulphate
SET	Su(var)3-9, E(z), Trx domain
SNF	sucrose non-fermenting
SNL	speckled nuclear localisation
snRNA	small nuclear RNA
SP1	specificity protein 1
ss	single stranded
SSC	saline sodium citrate
STAT	signal transducer and activator of transcription
SUMO	small ubiquitin-like modifier
SWI	switching defective
TAE	tris-acetate-EDTA
TAF	TBP associated factor
TAL1	T-cell acute lymphocytic leukaemia protein 1
TBE	tris-borate-EDTA
TBP	TATA binding protein

TBS-T	tris buffered saline–Tween20
TDG	thymine DNA glycosylase
TE	tris-EDTA
TEMED	tetra-methyl-ethylene-diamine
TET	ten-eleven translocation
TF	transcription factor
TIE2	*tunica interna* endothelial 2
tRNA	transfer RNA
TrxG	trithorax group
TSS	transcription start site
U	unit
UCSC	University of California, Santa Cruz
UV	ultraviolet
V	Volts
VEGF	vascular endothelial growth factor
WT	wild type

1. INTRODUCTION

1.1. The cell

The cell is the smallest living entity and has been referred to as “the building block of life”. Although the details of the evolutionary origin of the first cell are not entirely clear, it is obvious that all cells originate from pre-existing cells, through the process of mitosis (or meiosis for germ cells). The eukaryotic cell is surrounded by a two-layered lipid and protein membrane, the plasma membrane (Singer and Nicolson, 1972). Inside the plasma membrane lies the cytoplasm where all the cellular organelles are located (*Figure 1.1*).

The organelles are membranous structures with highly specific roles including energy conversion, metabolism and protein synthesis (Bainton, 1981; de Duve, 1996; McBride et al., 2006; Palade, 1975). The nucleus is the organelle where the blueprints for a cell are stored in the form of deoxy-ribonucleic acid (DNA) (Avery et al., 1979). It is surrounded by two double lipid membranes – the nuclear envelope – and is connected to the cytoplasm via the nuclear pores (reviewed in Lamond and Earnshaw, 1998).

A cell’s DNA contains functional units that encode for protein or ribonucleic acid (RNA) with specific functions, scattered between large pieces of DNA that do not (reviewed in Gall, 1981; Gregory, 2001). Such a coding unit is what is referred to as a gene. The DNA in the nucleus is

wrapped around small basic proteins – the histones (reviewed in Kornberg and Lorch, 1999) – forming a structure termed chromatin by W. Flemming in 1882.

The cell, although separated from its environment by the plasma membrane, is not isolated from it. It can dynamically respond to environmental signals received from the extracellular environment by receptor proteins which will initiate a signalling cascade that will eventually reach the nucleus and change the gene expression program of the cell in response to the environmental cue received (reviewed in Brivanlou and Darnell, 2002).

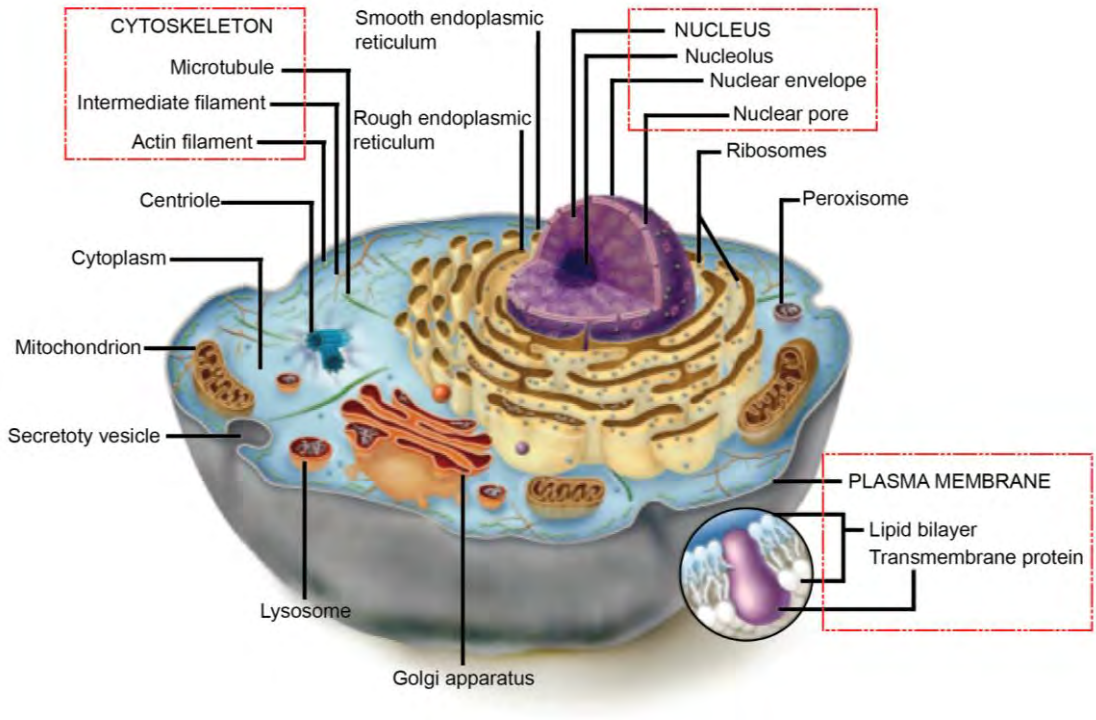

Figure 1.1. Schematic representation of an animal cell. Figure created by Tilottama Chatterjee, publically available on NotesMaster (http://www.notesmaster.com/notes/syllabus/viewer/3099-plant-animal-and-microbe-cells).

1.2. Chromatin and transcription

1.2.1. Chromatin structure

Chromatin is the physiological substrate of all DNA templated processes, such as DNA replication, DNA repair and of course transcription of messenger RNA (mRNA). Chromatin is mainly composed of DNA and a collection of proteins (Kornberg, 1974). Two copies of each of the four core histone proteins – H2A, H2B, H3 and H4 – form an octamer around which approximately 147bp of DNA are wound (Luger et al., 1997) (*Figure 1.2*). The histones are small, highly basic proteins and contain a globular "histone fold" domain and an N-terminal unstructured tail. The first role that histones have to play is compacting the approximately 1.5 meters of DNA (in humans) in the confined space (1-10μm in diameter) of the nucleus. A first degree of compaction is achieved by winding DNA approximately 1.7 times around the histone octamer, forming a nucleosome (a schematic representation of the different degrees of chromatin compaction is presented in *Figure 1.3*). A series of nucleosomes is the 10nm wide "beads-on-a-string" structure observed by electron microscopy in conditions of low ionic strength (Olins and Olins, 1974), which was thought to be the native form of chromatin in the nucleus. Later studies have demonstrated the existence of a higher order chromatin structure, the 30nm fibre, observed at ionic strengths closer to physiological levels and whose formation requires the presence

of the linker histone H1 (reviewed in Felsenfeld and McGhee, 1986). The existence and the exact structure of the 30nm fibre and the position of H1 in this structure *in vivo* is still a matter of intense debate (Eltsov et al., 2008; Grigoryev and Woodcock, 2012; Robinson and Rhodes, 2006). During mitosis, just before metaphase, all nuclear DNA is highly compacted to form the “metaphasic chromosomes” which are eventually split between the two daughter cells at the end of mitosis.

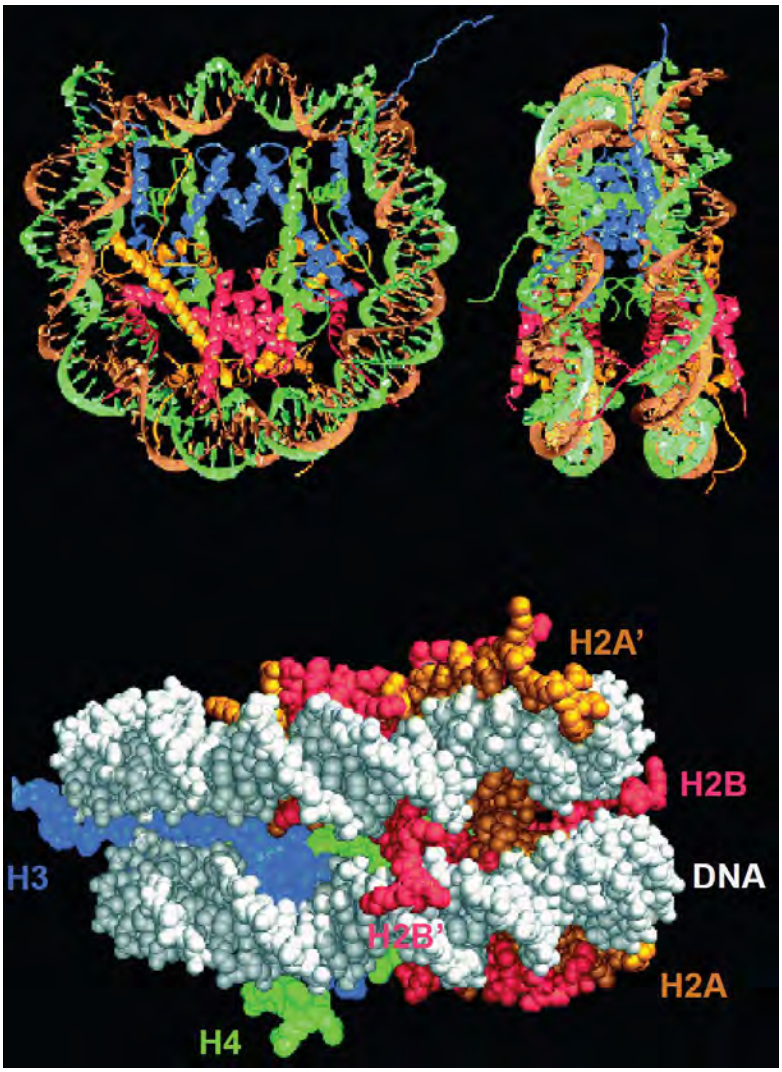

Figure 2.2. Crystal structure of the nucleosome (Figure from Luger et al., 1997). **Top panel:** Top (left) and side (right) view of the nucleosome particle. DNA backbones in brown and turquoise, H3: blue, H4: green, H2A: yellow, H2B: red. **Bottom panel:** The N-terminal tails protrude outside the core nucleosome particle

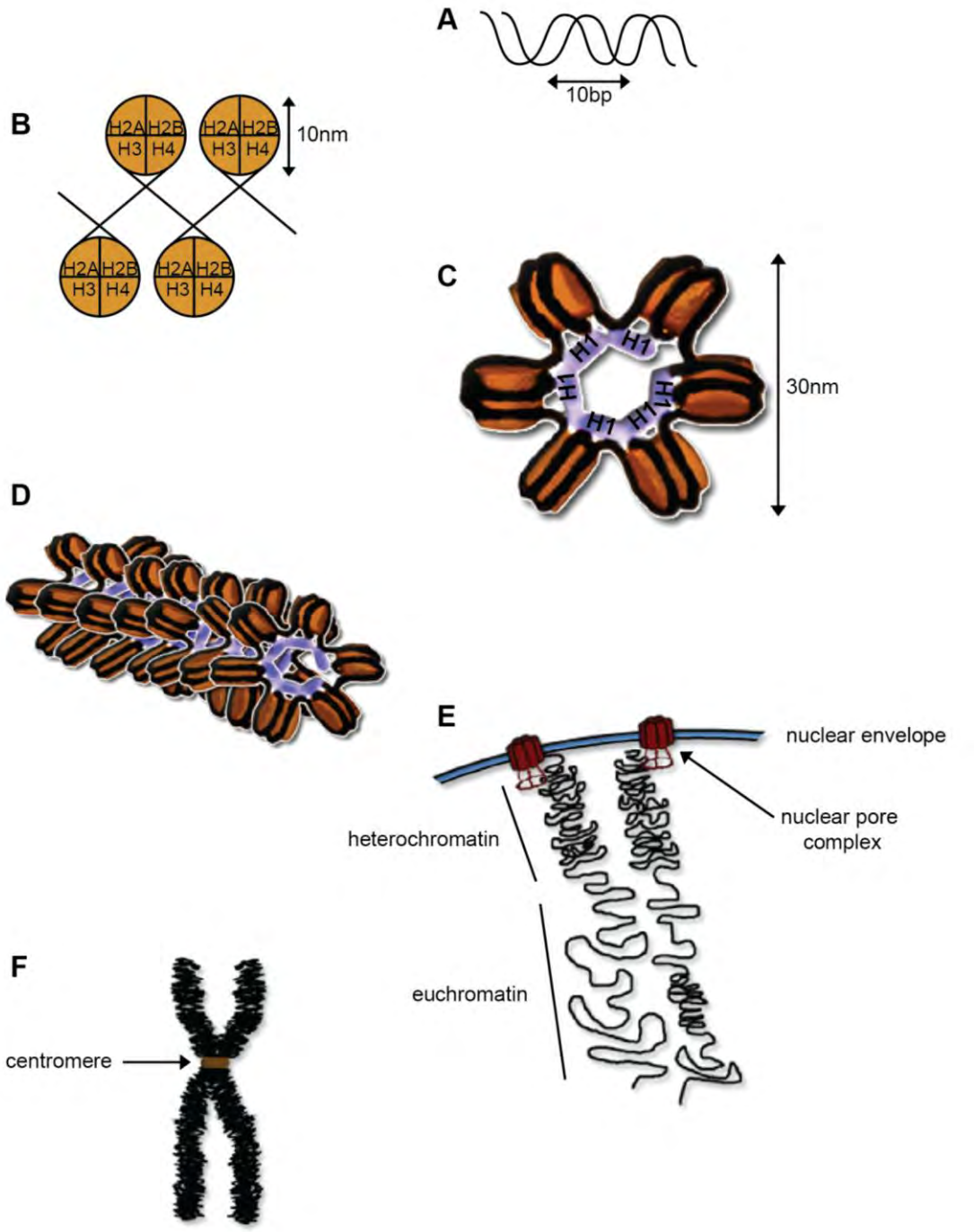

Figure 1.3. Chromatin structure: from the DNA double helix to metaphasic chromosomes. **A:** The DNA double helix, two complementary strands forming a helical structure with a 10bp periodicity. **B:** DNA is wrapped 1.7 times around a histone octamer, made up of two copies of each of the core histones. **C:** Addition of histone H1 (speculative position of H1 illustrated in blue) facilitates the formation of the 30nm fibre. **D:** the 30nm solenoid fibre. **E:** Chromatin is further compacted in the nucleus. Centromeric and telomeric regions reside in heterochromatin, a highly compacted structure, while active regions reside mainly in the less compacted euchromatin. **F:** A highly compacted metaphasic chromosome, two copies of the same DNA molecule joined at the centromere.

1.2.2. Basal transcription

The process of transcription of DNA into RNA is a multi-step enzymatic reaction and involves proteins binding to DNA, unwinding of the double helix and polymerization of ribonucleotides to form a RNA molecule of complementary sequence to the DNA strand from which it originated. RNA transcription is mediated by protein complexes comprised of general transcription factors and a catalytic RNA polymerase subunit.

1.2.2.1. RNA polymerases

There are three main RNA polymerases in all eukaryotes and all of them function in large complexes with general transcription factors. The three eukaryotic RNA polymerases share extensive structural similarity between them and also with the prokaryotic RNA polymerase (Allison et al., 1985). RNA polymerase I is located in the nucleolus (Roeder and Rutter, 1969) and is responsible for transcribing the 18S and 28S ribosomal RNAs (rRNAs) (Zylber and Penman, 1971). RNA polymerase II transcribes the regulatory miRNA species (Lee et al., 2004), protein coding genes into mRNAs (Weil et al., 1979) and possibly other regulatory non-coding RNA (ncRNA) species. RNA polymerase III is located in the nucleoplasm (Roeder and Rutter, 1969) and generates the transfer RNAs (tRNAs) and small nuclear RNAs (snRNAs) (Hall et al., 1982).

1.2.2.2. The RNA polymerase II core promoter

In order for mRNA transcription to take place, the transcriptional machinery first needs to assemble on a specific DNA sequence. This "docking site" is provided by the core promoter, which assembles all the necessary factors for transcription to start. Core promoters are absolutely required for the initiation of mRNA transcription and they are located in the immediate vicinity of the transcription start site (TSS).

The RNA polymerase II core promoter contains a set of conserved DNA sequence elements, including the initiator sequence (Inr), the TATA box, the upstream and downstream $TF_{II}B$ recognition elements (BREu/d), the downstream promoter element (DPE) and the motif ten element (MTE) (for a comprehensive review on all core promoter elements see Juven-Gershon et al., 2008; Juven-Gershon and Kadonaga, 2010; Smale and Kadonaga, 2003). A schematic of a core promoter is shown in *Figure 1.4.* The Inr spans from -3bp to +5bp, thus encompassing the TSS (Smale and Baltimore, 1989) and the TATA Box is located approximately 25bp upstream of the transcription start site (TSS). The $TF_{II}D$ complex can bind on both of these elements via the TBP (TATA-box) and TAF1 and TAF2 (Inr) subunits (Benoist and Chambon, 1981; Dierks et al., 1983; Juven-Gershon and Kadonaga, 2010; Mathis and Chambon, 1981). The BREu and BREd have been found immediately up/downstream of the TATA-box, on a subset of TATA-box containing promoters and can either increase or decrease transcription rates (Deng and Roberts, 2005; Evans et al., 2001;

Lagrange et al., 1998). The DPE is located around +28bp to +32bp mostly in promoters lacking a TATA box (Burke and Kadonaga, 1996). On those promoters $TF_{II}D$ binds to the Inr and DPE sequences. Spacing between those two sequence elements appears to be critical for the function of promoters containing such motifs (Kutach and Kadonaga, 2000). The MTE was found immediately upstream of the DPE and is thought to be a binding site for $TF_{II}D$ (Lim et al., 2004; Ohler et al., 2002). The MTE functions together with the Inr but can be independent of the TATA-box and the DPE (Lim et al., 2004). These elements provide extensive diversity in core promoter composition and it has been shown that certain activators preferentially interact with core promoters harbouring different combinations of these elements (reviewed in Juven-Gershon and Kadonaga, 2010). The diversification of core promoters is further enhanced by the presence of TBP-related factors and atypical TAFs that assemble into different complexes and preferentially target specific core promoters in different cell types / developmental stages (reviewed in Goodrich and Tjian, 2010; Juven-Gershon and Kadonaga, 2010).

Recent genome-wide studies (reviewed in Lenhard et al., 2012) have changed our view of core promoters from TATA-box versus TATA-less promoters. Promoters are now classified in three different groups, Type I, II and III promoters. Type I promoters closely resemble the core promoters originally described in *Drosophila*. These promoters usually contain a TATA-box that defines a unique TSS and normally do not overlap a CpG island (Akalin et al., 2009; Carninci et al., 2006; Lenhard et al., 2012; Rach

et al., 2011; Yamashita et al., 2005). Type I promoters are generally associated with highly inducible/regulated genes (Lenhard et al., 2012). Type II promoters are generally found on housekeeping genes and are TATA-depleted. These promoters have numerous transcription start sites over a broader region and overlap a CpG island. They often contain sequence elements that are not positionally fixed, (Motif 1 or 6, DRE) and are usually found associated with constitutively expressed genes (Akalin et al., 2009; Carninci et al., 2006; Lenhard et al., 2012; Rach et al., 2011; Yamashita et al., 2005). Type III promoters are also TATA-depleted and generally overlap multiple large CpG islands that often extend well into the coding region. Initiation from these promoters happens from a narrower region than on Type II, however they do not have the very well defined TSS of Type I promoters. Type III promoters are generally associated with developmentally regulated genes and are often targets for Polycomb-mediated repression (Engstrom et al., 2007; Lenhard et al., 2012).

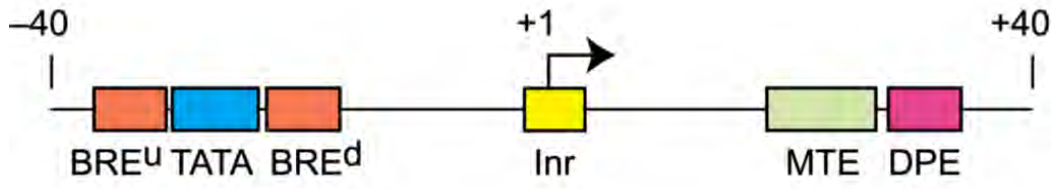

Figure 1.4. Structure of a RNA polymerase II core promoter (Figure adapted from Juven-Gershon and Kadonaga, 2010). BREu: upstream $TF_{II}B$ recognition element, TATA: TATA-box, BREd: downstream $TF_{II}B$ recognition element, Arrow: transcription start site, Inr: initiator sequence, MTE: motif ten element, DPE: downstream promoter element.

1.2.3. Sequence specific transcription factors

Promoters also contain binding sites for other – non-general – sequence specific transcription factors (TFs) (Bram and Kornberg, 1985; Cordingley and Hager, 1988; Lee et al., 1987a). TFs were first discovered in the 1980s (Bohmann et al., 1987; Dynan and Tjian, 1983; Lee et al., 1987a) and have been implicated in a variety of transcriptional processes and cellular responses (Baxter et al., 1972; Berleth et al., 1988; Driever and Nusslein-Volhard, 1988; Hager and Palmiter, 1981; Nakabeppu et al., 1988). These TFs commonly contain a DNA binding domain and a *trans*-activation domain (Ptashne and Gann, 1997; Triezenberg, 1995) (*Figure 1.5*), which can recruit histone modifying activities, chromatin remodelers or the basal transcription machinery (reviewed in Lee and Young, 2000). TFs often work in collaboration to establish the gene expression programs that shape cellular identity (among others Krysinska et al., 2007; Nottingham et al., 2007; Rodriguez et al., 2005).

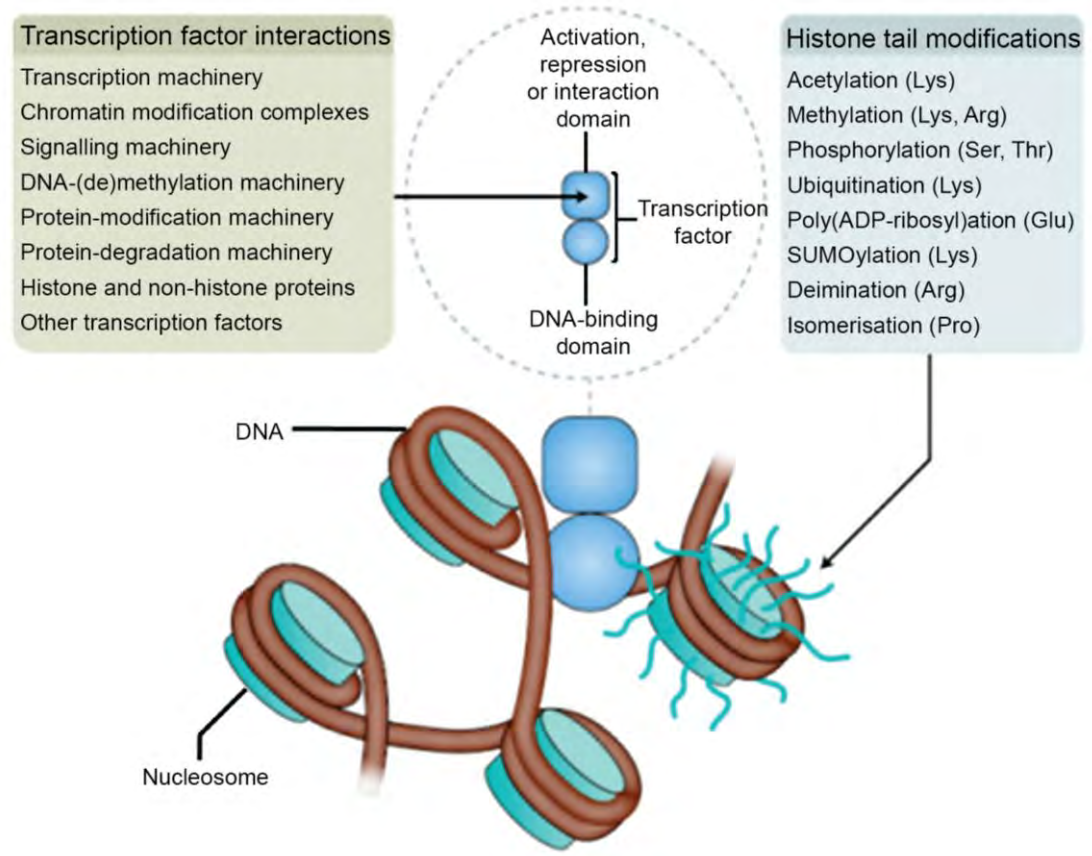

Figure 1.5. Transcription factors bind DNA in the context of chromatin (Figure from Bonifer and Bowen, 2010). They exert their function through interaction with a variety of other protein factors. The nucleosome on the right is illustrated with its N-terminal histone tails, which are targets for extensive post-translational modification.

1.2.4. Enhancer and insulator elements

Promoters are not the only *cis*-regulatory elements found in eukaryotic genomes (Bergman et al., 1984; Thanos et al., 1988). Enhancers also play an important role in regulation of gene expression (Moreau et al., 1981; Wasylyk and Chambon, 1983). These are DNA sequences which can increase the transcriptional activity of a target promoter. They bind activating transcription factors (TFs) (Brady and Khoury, 1985; Thanos et al., 1988) and can act as signal integration hubs (Lee et al., 1987a; Lee et

al., 1987b). Although enhancers can be located megabases away from the TSS on the one-dimensional DNA sequence, they are thought to be in very close proximity in the three-dimensional space of the nucleus when they are active (Heuchel et al., 1989; Mueller-Storm et al., 1989; Su et al., 1991). Physical interactions between an enhancer and a target promoter are established by the cooperative action of TFs that bind the enhancer and promoter (*Figure 1.6A*) (Su et al., 1991). In some cases these interacting TFs and other protein complexes form a tight complex that constitutes the so-called enhanceosome (Thanos and Maniatis, 1995). Enhancers act by significantly enhancing the rate of transcription (Moreau et al., 1981; Wasylyk and Chambon, 1983). This is achieved by increased recruitment of RNA polymerase II to the promoter (Treisman and Maniatis, 1985; Weber and Schaffner, 1985) or stabilisation of the RNA polymerase II complex (Walters et al., 1995; Walters et al., 1996). Enhancers can provide tissue specificity in gene expression (Edlund et al., 1985; Ephrussi et al., 1985; Gillies et al., 1983), by recruiting tissue-specific TFs (Staudt et al., 1986). In both cases, enhancers may interact with the mediator complex that can recruit the basal transcription machinery (reviewed in Malik and Roeder, 2010) and / or histone modifying complexes that set up a chromatin environment that is permissive for transcription (Merika et al., 1998; Vernimmen et al., 2011).

Since enhancers can affect transcription over long distances, it is imperative that their actions are targeted to the correct promoter. This is directed by another type of *cis*-regulatory elements, insulators. These are DNA sequences that are recognised and bound by specific insulator proteins – mainly CTCF – and can mediate DNA loop formation to promote or prevent enhancer-promoter interactions (Bell et al., 1999) (*Figure 1.6B*). Insulators are also involved in preserving heterochromatin boundaries (Saitoh et al., 2000) (*Figure 1.6C*).

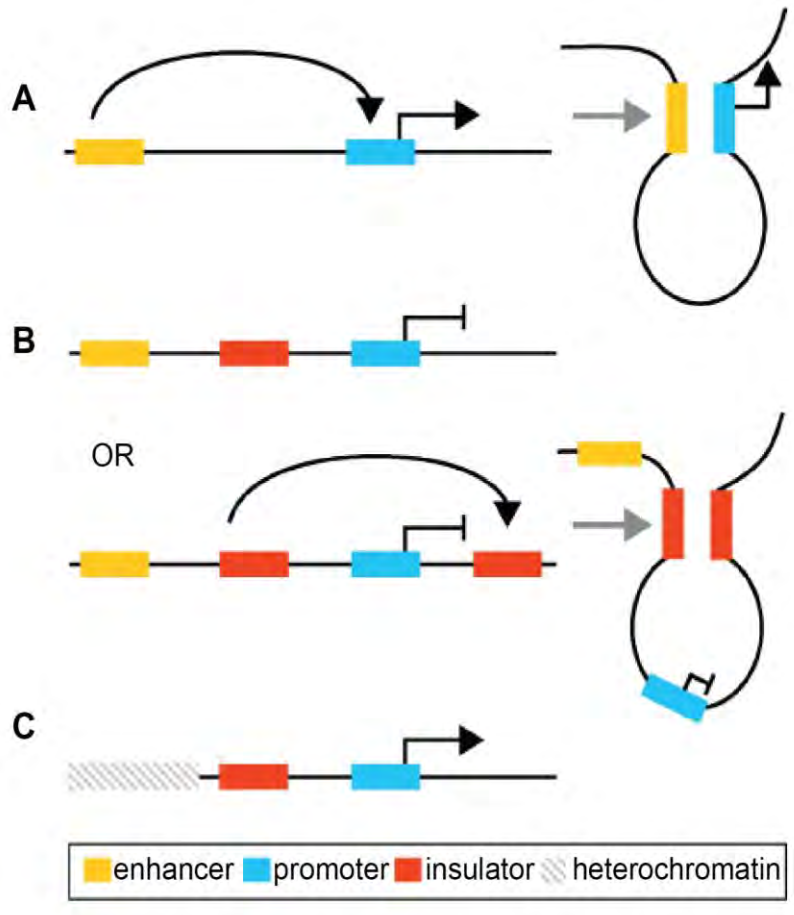

Figure 1.6. Co-operative function of *cis* regulatory elements (Figure from Krivega and Dean, 2012) **A:** Enhancers are brought in close proximity to their target promoters through the action of sequence specific TFs. **B:** Insulators function by blocking enhancer-promoter interactions, possibly by generating loops, thereby targeting an enhancer to a specific promoter. **C:** Insulator elements also function by presenting a "road-block" to the expansion of heterochromatin domains

1.2.5. Chromatin remodelling

The higher order chromatin structures described previously are by no means static. Chromatin associated proteins interact only transiently with the chromatin fibre and are rapidly replaced (Meshorer et al., 2006; Phair et al., 2004a; Phair et al., 2004b). When a cell receives differentiation, stress or other signals it will need to change its expression profile. The local chromatin environment at promoters that need to be activated or silenced needs to be altered accordingly.

If a nucleosome is situated over the transcription start site, the basal transcription machinery cannot assemble on the promoter (Workman and Roeder, 1987). A first step towards gene activation would then be displacement of that nucleosome (Fletcher et al., 2002). This is achieved by ATP-dependent nucleosome remodelling complexes that are directly or indirectly recruited to target promoters by sequence specific transcription factors (Belikov et al., 2004; Fryer and Archer, 1998; Truss et al., 1995; Tsukiyama et al., 1994; Yie et al., 1999). These remodelers utilize the energy provided by the hydrolysis of ATP to slide nucleosomes to different positions, displace H2A/H2B dimers leaving the core H3/H4 tetramer intact, evict a whole nucleosome or mediate exchange of the core histones for histone variants (reviewed in Kouzarides, 2007; Segal and Widom, 2009; Talbert and Henikoff, 2010; Workman and Kingston, 1998).

The activating action of ATP-dependent remodelers is based on the fact that they can either generate a nucleosome free region over a gene's promoter or loosen the histone-DNA contacts sufficiently to allow binding of other effector proteins and the basal transcription machinery (Agalioti et al., 2000; Almer et al., 1986; Archer et al., 1991). Regardless of whether the nucleosome is displaced or disrupted, the presence of the remodelling complex and binding of transcription factors increases DNA accessibility. This can be detected experimentally as increased sensitivity to endonucleases such as DNaseI (McGhee et al., 1981; Weiss et al., 1986).

The ATP-dependent remodelers are categorised in four families: SWI/SNF (switching defective / sucrose non-fermenting), ISWI (imitation SWI), NuRD (nucleosome remodelling and deacetylation) and INO80 (inositol requiring 80) (reviewed in Cairns, 2005, 2007; Clapier and Cairns, 2009). These ATP-dependent remodelers are contained within larger multi-subunit complexes that have distinct roles in gene regulation.

SWI/SNF complexes act mostly in gene activation (reviewed in Martens and Winston, 2003) by generating irregular nucleosomal arrays that leave binding sites for transcription factors exposed and available for binding.

ISWI complexes generally generate regularly spaced nucleosomal arrays that result in gene repression (reviewed in Corona and Tamkun, 2004), although there have been examples of ISWI mediated activation (reviewed in Badenhorst et al., 2002; Morillon et al., 2003). NuRD complexes are generally repressive but can act to activate transcription of rRNA species (reviewed in Eberharter and Becker, 2004).

INO80 also has both repressive and activating functions. It has also been implicated in DNA repair via interaction with the histone H2A variant, H2A.X and in distribution of another histone H2A variant, H2A.Z at active promoters (Jonsson et al., 2004; Kobor et al., 2004; Mizuguchi et al., 2004; Morrison et al., 2004; van Attikum et al., 2004). Notably, nucleosomes adjacent to active promoters also contain a histone H3 variant, H3.3 (Chow et al., 2005; Mito et al., 2005; Schwartz and Ahmad, 2005). The combination of H3.3 and H2A.Z has been reported to form very unstable nucleosomes and exclude histone H1 (Braunschweig et al., 2009; Jin and Felsenfeld, 2007; Jin et al., 2009), thus contributing to DNA accessibility and transcription (reviewed in Henikoff, 2009).

1.2.6. Histone post-translational modifications

The histone proteins – the unstructured N-terminal tails in particular – are targets for extensive post-translational modification (PTM) (Allfrey et al., 1964; Bannister and Kouzarides, 2011). These PTMs have been proposed to have regulatory roles in most DNA related processes and most notably transcription, in which context they have been studied the most (some of the known histone PTMs are illustrated in *Figure 1.7*).

The histone tails do not affect the structure of individual nucleosome core particles, but they may impact on the overall chromatin environment and structure by recruiting other protein co-factors, mediating inter-

nucleosomal contacts and altering histone-DNA binding affinity (reviewed in Cockerill, 2011; Peterson and Laniel, 2004).

The presence of different modifications at different target sites gave rise to speculation on the existence of a “histone code” (Jenuwein and Allis, 2001; Strahl and Allis, 2000). The histone code hypothesis has been a matter of intense debate within the scientific community but it is still unclear to which extent histone PTMs may have an instructive role in gene transcription. The main reason for this is that the few studies that hint to such mechanisms are mostly correlative and do not provide direct evidence of causality (reviewed in Henikoff and Shilatifard, 2011; Turner, 2012). Moreover, extensive redundancy within different histone modifying enzyme families complicates things even further and makes genetic-manipulation approaches difficult to interpret. Though undoubtedly histone PTM play some role in transcriptional regulation in mammals – as illustrated by the severity of phenotypes associated with deletions of histone modifying enzymes – there is little evidence of direct causality (Bungard et al., 2010; Cao and Zhang, 2004; Smith et al., 2011). An example showing that histone modifications may be a by-product of other processes comes from murine ES cells (reviewed in Turner, 2012). In these cells G9a methylates lysine 9 on histone H3 on the promoter of the pluripotency gene *Oct4* upon differentiation (Feldman et al., 2006), followed by epigenetic gene silencing. The mechanism was thought to involve recruitment of heterochromatin protein 1 (HP1) via the H3K9 methyl mark, which in turn would recruit the DNA methyltransferases

DNMT3a/b. A more recent study however, has shown that this silencing process in independent of G9a catalytic activity (Epsztejn-Litman et al., 2008). In contrast, the same study shows that although methylation of lysine 9 on histone H3 was not required for DNA methylation, it was required for HP1 association and chromatin compaction. Further studies in yeast have shown that the main role of histone H3 lysine 9 methylation may be to recruit HP1 which can then "bridge" bucleosomes together and result in chromatin compaction (reviewed in Turner, 2012). These results reinforce the notion that histone modifications are part of a cooperative process where the interacting partners and their precise roles are only incompletely understood. Additionally, the use of the word "code" presumes that combinations of histone PTMs would have a biological output different to the individual PTMs added together. This however, has not been observed to date by any of the numerous (among many others Hoskins et al., 2011; Kharchenko et al., 2011; Mikkelsen et al., 2007) genome-wide histone PTM mapping studies. A list of different histone PTMs, the enzymes that catalyse them and their attributed transcriptional function is shown in *Table 1.1.*

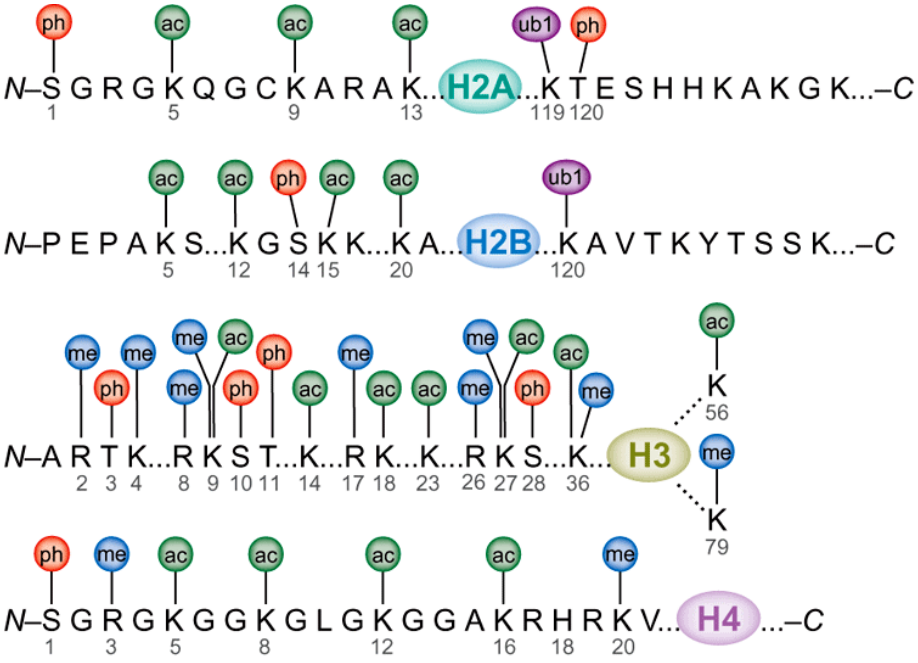

Figure 1.7. Schematic of histone post-translational modifications (Bhaumik et al., 2007). ac: acetylation, me: methylation, ph: phosphorylation, ub1: mono-ubiquitination.

1.2.6.1. Histone phosphorylation, ubiquitination, SUMOylation and proline-isomerisation

Phosphorylation of histones takes place on serine, threonine and tyrosine residues and is a highly dynamic modification. Histone phosphorylation is regulated by a balance of kinases and phosphatases. Histone phosphorylation has been implicated in chromatin condensation during mitosis (Gurley et al., 1974) and signal-dependent activation of inducible genes (Clayton et al., 2000; Soloaga et al., 2003). Notably, phosphorylation confers a negative charge to the histone molecule which could diminish the electrostatic interactions between the nucleosome core and the negatively charged DNA backbone.

Histone ubiquitination, although observed as early as 1982 (Levinger and Varshavsky, 1982), has not yet been attributed a clear role in gene transcription, with the exception of mono-ubiquitination of histones H2A and H2B at lysines 119 and 120 that have been clearly implicated in gene repression and activation respectively (Wright et al., 2012).

The role of SUMOylation (small ubiquitin-like modification) is even more unclear than that of ubiquitin. SUMOylation appears to act by preventing ubiquitination and acetylation of the same residues (Iniguez-Lluhi, 2006; Long et al., 2005; Wotton and Merrill, 2007).

Proline isomerisation, although not a covalent PTM, can quite drastically change the structure of a histone tail and affect transcription by inter-converting proline residues between the *cis* and *trans* conformation (Chen et al., 2006; Nelson et al., 2006).

1.2.6.2. Histone acetylation

Acetylation was one of the first observed histone modifications (Allfrey et al., 1964). The discovery and characterisation of the first histone acetyltransferase (HAT) came many years later (Brownell and Allis, 1995; Kleff et al., 1995). The association of histone acetylation with active transcription was hypothesised long before the discovery of HATs (Clever and Ellgaard, 1970; Gallwitz and Sekeris, 1969; Pogo et al., 1966; Wilhelm and McCarty, 1970).

A large number of HATs have been characterised since then and they can be classified in two categories. Type B HATs are cytoplasmic and acetylate individual histones before they are translocated into the nucleus. This aids the assembly of the nucleosome core particle, after which the marks laid down by type B HATs are replaced by a locus appropriate pattern (Parthun, 2007; Scharf et al., 2009).

Type A HATs reside in the nucleus and belong to one of three protein families: MYST, GNAT and CBP/p300. HATs target a plethora of lysine residues on all the core histones with rather low substrate specificity (reviewed in Bannister and Kouzarides, 2011).

The effect of lysine acetylation is loss of the lysine's positive charge. Given the number of lysines that can be acetylated in any one nucleosome – one on H2A, four on H2B, five on H3 and four on H4 – the net charge effect can be quite significant. Early experiments with the HDAC inhibitor sodium butyrate demonstrated that histone hyper-acetylation results in increased chromatin accessibility (Simpson, 1978; Vidali et al., 1978) due to small attenuation of the electrostatic interactions between the nucleosome core particle and the DNA backbone (Mathis and Chambon, 1981) and interference with intra-nucleosomal contacts, causing a general de-compaction (Oliva et al., 1990).

One particular acetylation site however, has greater implications than charge neutralisation alone. A more recent study in yeast demonstrated that acetylation of lysine 16 on histone H4 along with eviction of histone H1 is sufficient to de-compact the 30nm chromatin fibre (Robinson et al.,

2008). This could provide a very elegant and simple mechanism for local chromatin de-compaction to allow, for example, passage of a transcribing RNA polymerase.

Moreover, lysine acetylation is recognised by bromodomains (Owen et al., 2000) and can thus serve as a tethering point for proteins that harbour such domains. Many of the ATP-dependent remodelling complexes mentioned previously contain proteins with such domains (Kasten et al., 2004; Tamkun et al., 1992), providing a mechanism by which a remodelling complex could be targeted to nucleosomes on active regions of the genome.

1.2.6.3. Histone methylation

Histone methylation was discovered in the same study as histone acetylation (Allfrey et al., 1964). Its role however had remained elusive for over 20 years. Histones are methylated on lysine and arginine residues. Lysines can be mono-, di-, or tri-methylated, while arginines can be mono- and symmetrically or asymmetrically di-methylated.

The functional outcome of lysine methylation depends on the position of the modified lysine. The first identified histone lysine methyltransferase was suppressor of variegation 3-9 homologue 1 (SUV39H1) (Rea et al., 2000). This enzyme has specific activity against lysine 9 of histone H3. This modification has been implicated in transcriptional silencing,

heterochromatin formation and maintenance, DNA methylation (will be discussed later) and genome stability (reviewed in Shinkai, 2007). Transcriptional repression in this case is mediated by the recruitment of the DNA methylation machinery and heterochromatin protein 1 (HP1) to the H3K9me mark (Fuks et al., 2003a; Vaute et al., 2002).

Another repressive lysine methyl-mark is deposited on lysine 27 of histone H3 by proteins of the polycomb group (Cao et al., 2002). The polycomb repressive complex 2 (PRC2) is recruited to target genes by a mechanism that involves non-coding RNA species (Kaneko et al., 2010; Kanhere et al., 2010). Its catalytic subunit – enhancer of zeste homologue 2 (EZH2) – deposits methyl-moieties on lysine 27 of histone H3 that subsequently serve to recruit the polycomb repressive complex 1 (PRC1). PRC1 mediates mono-ubiquitination of lysine 119 of histone H2A, a mark associated with transcriptional repression (Wang et al., 2004a).

Methylation of histone H3 on lysine 4 has been associated with active promoters and enhancers. Lysine 4 tri-methylation is found on active promoters, as demonstrated by many genome wide mapping studies of this mark. Mono-methylation of the same residue can be observed on enhancers together with acetylation of lysine 27 of histone H3 (Heintzman et al., 2007). Additionally, lysine 4 tri-methylation has also been found on CpG islands that lack promoter activity (Thomson et al., 2010) and has been proposed to confer protection from *de novo* DNA methylation (Ooi et al., 2007).

Histone methylation does not change the charge of the histone tail and is a rather stable modification. In fact it is so stable that it was thought to be permanent. The discovery of methyl-arginine deimination provided a mechanism by which the methyl-arginine could be removed by conversion into citrullin (Cuthbert et al., 2004; Wang et al., 2004b). In the same year the first lysine specific demethylase (LSD1) was discovered. LSD1 can demethylate mono- and di-methylated lysines (Shi et al., 2004). Following these discoveries, a second family of demethylases was discovered (Tsukada et al., 2006). These proteins contain a catalytic jumonji domain and are able to demethylate all forms of methylated lysines, via a different chemical reaction to the one catalysed by LSD1.

Table 1.1 Histone modifications associated with transcription (adapted from Li et al., 2007)

Modification	Position	Enzyme				Function in transcription
		S. Cerevisiae	*S. Pombe*	*Drosophila*	Mammals	
K Methylation	**H3** K4	Set1	Set1	Trx, Ash1	MLL, ALL-1 Set9/7 ALR-1/2 ALR, Set1	Activation
	K9	n/a	Clr4	Su(var)3-9 Ash1	Suv39h, G9a Eu-HMTase I ESET, SETBD1	Repression, activation
	K27				E(Z)	Repression,
	K36	Set2		HYPB Smyd2 NSD1		Rpd3S mediated repression
	K79	Dot1		Dot1L		Activation
	H4 K20		Set9	PR-Set7 Ash1	PR-Set7 SET8	Silencing
R Methylation	**H3** R2				CARM1	Activation
	R17				CARM1	Activation
	R26				CARM1	Activation
	H4 R3				PRMT1	Activation
Phosphorylation	**H3** S10	Snf1				Activation
Ubiquitination	**H2B** K120/123	Rad6, Bre1	Rad6		UbcH6 RNF20/40	Activation
	H2A K119				hPRC1L	Repression
Acetylation	**H3** K56					Activation
	H4 K16	Sas2, NuA4		dMOF	hMOF	Activation
	Htz1 K14	NuA4, SAGA				Activation

1.2.6.4. Crosstalk between histone PTMs

The plethora of different histone modifications inspired the "histone code" hypothesis discussed earlier. Though the existence of a "code" *per se* was never verified, certain elements of that theory hold true (reviewed in Turner, 2012). There have been numerous studies reporting cross-talk between different histone PTMs in a way that a histone PTM on one histone tail may induce or preclude specific modification of different residues on the same (*in cis*) or a different histone (*in trans*) tail (reviewed in Bannister and Kouzarides, 2011). A summary of interdependent histone PTMs is shown schematically in *Figure 1.8*.

A good example of histone PTM cross-talk *in cis* was described in *Saccharomyces cerevisiae*. The Snf1 kinase phosphorylates histone H3 on serine 10. This modification recruits the Gcn5 acetyltransferase which in turn acetylates lysine 14 of histone H3. Members of the 14-3-3 protein family are then recruited to the doubly modified histone H3 tail and mediate transcriptional activation (Walter et al., 2008).

The connection between serine 10 phosphorylation and lysine 14 acetylation was also demonstrated in humans (Clayton et al., 2000; Mateescu et al., 2004). The study by Mateescu et al. (2004) revealed that this double phospho-acetylation mark negates the effects of lysine 9 methylation by displacing HP1 – the main downstream effector of lysine 9 methylation.

Serine 10 phosphorylation is also involved in *trans*-histone cross-talk. A recent study (Zippo et al., 2009) demonstrated that the kinase PIM1 is recruited to the H3 lysine 9 acetylated nucleosomes on the promoter of the FOSL1 gene and phosphorylates serine 10 of histone H3. This results in acetylation of lysine 14 on H3 and the recruitment of 14-3-3 proteins. Those in turn recruit the MOF acetyltransferase that acetylates lysine 16 of histone H4. This chromatin environment is favourable for recruitment of the bromodomain protein BRD4 which can recruit the positive transcription elongation factor b (p-TEFb) to promote transcriptional elongation (Hargreaves et al., 2009; Yang et al., 2008; Zippo et al., 2009).

Another example of PTM cross-talk *in trans* is that between mono-ubiquitination of H2B and methylation of lysine 4 and lysine 79 on H3 – both being hallmarks of transcriptionally active genes. In *S. cerevisiae*, the protein Cps35 binds to mono-ubiquitinated histone H2B and recruits a protein complex that contains the Set1 histone methyltransferase to mediate tri-methylation of lysine 4 of histone H3 (Lee et al., 2007). An analogous phenomenon was also observed in human cells (Kim et al., 2009). Methylation of lysine 79 of histone H3 is mediated by the enzyme Dot1 which also binds to the H2B mono-ubiquitin (McGinty et al., 2008; Weake and Workman, 2008).

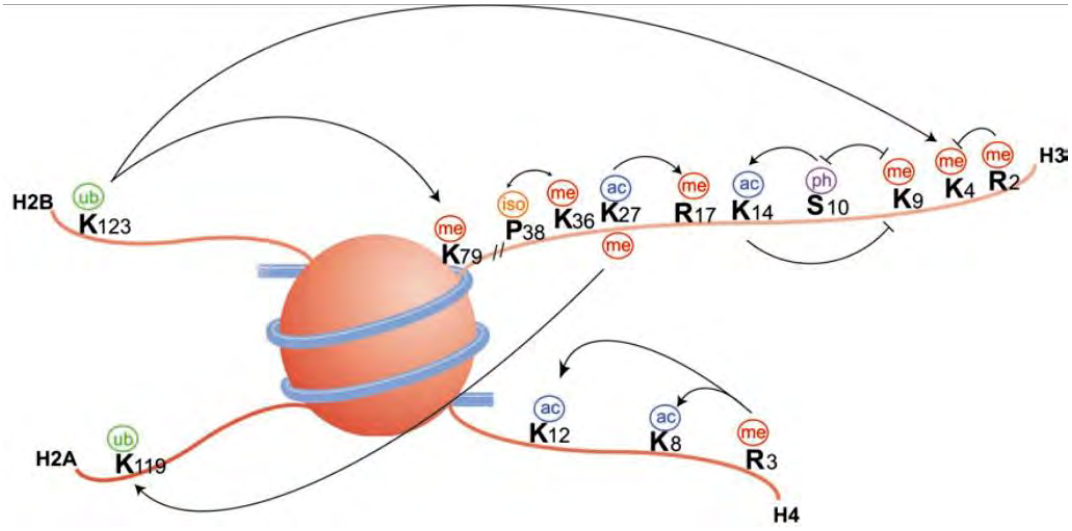

Figure 1.8. Schematic of histone PTM cross-talk (figure from Bannister and Kouzarides, 2011). ac: acetylation, me: methylation, ph: phosphorylation, ub: mono-ubiquitination, iso: proline isomerisation. Arrow heads indicate a positive effect while blocked lines indicate a negative effect.

1.2.7. Transcription by RNA polymerase II

1.2.7.1. Transcription initiation

For RNA polymerase II mediated transcription, the catalytic subunit needs to be recruited to the core promoter and positioned just upstream of the TSS. The RNA polymerase II holoenzyme is bound to the promoter by the $TF_{II}D$ protein complex (*Figure 1.9*). This complex contains the TATA binding protein (TBP) and several TBP associated factors (TAFs).

After $TF_{II}D$ binding on the promoter, $TF_{II}A$ is recruited, which in turn recruits $TF_{II}B$. $TF_{II}B$ interacts with Rpb1 – the RNA polymerase II catalytic subunit – via three different domains (Nikolov and Burley, 1997). However the presence of $TF_{II}B$ is not sufficient to recruit Rpb1. It is only after $TF_{II}F$ and $TF_{II}E$ are recruited to the complex that the main polymerase subunit is recruited. At this point the initiation complex is assembled and the RNA polymerase starts transcribing but terminates after a short distance.

After the incorporation of factors $TF_{II}H$ (ATPase, helicase and kinase activity) and $TF_{II}J$ the enzyme can start genuine transcription (Goodrich and Tjian, 1994). $TF_{II}H$ is responsible for the unwinding of the DNA double helix right in front of the polymerase to allow access to the template strand (Douziech et al., 2000) and also phosphorylates the C-terminal domain (CTD) heptapeptide (YSPTSPS) repeats of the enzyme at serine 5 (Hirose and Ohkuma, 2007).

This model was established a long time ago (reviewed in Buratowski, 1994) and has since been challenged on many different levels. As mentioned previously, there is a collection of different TBP related factors and TAFs that can assemble into atypical TF_{II} complexes that function on distinct promoter subsets. This suggests that the so-called general transcription factors may not be so general after all and that there is at least some transcriptional regulation at the level of the core promoter. Additionally, recent work (Apostolou and Thanos, 2008; Eskiw and Fraser, 2011; Osborne et al., 2004) has demonstrated that promoters are recruited to the RNA polymerase II and associated factors rather than the

factors to the promoter. These studies propose the existence of “transcription factories”, areas in the nucleus where there is increased concentration of RNA polymerase II, general transcription factors and even tissue specific transcription factors. Upon stimulus, genes that are to become activated are then thought to be recruited to these factories where all the necessary factors are present and transcription can initiate.

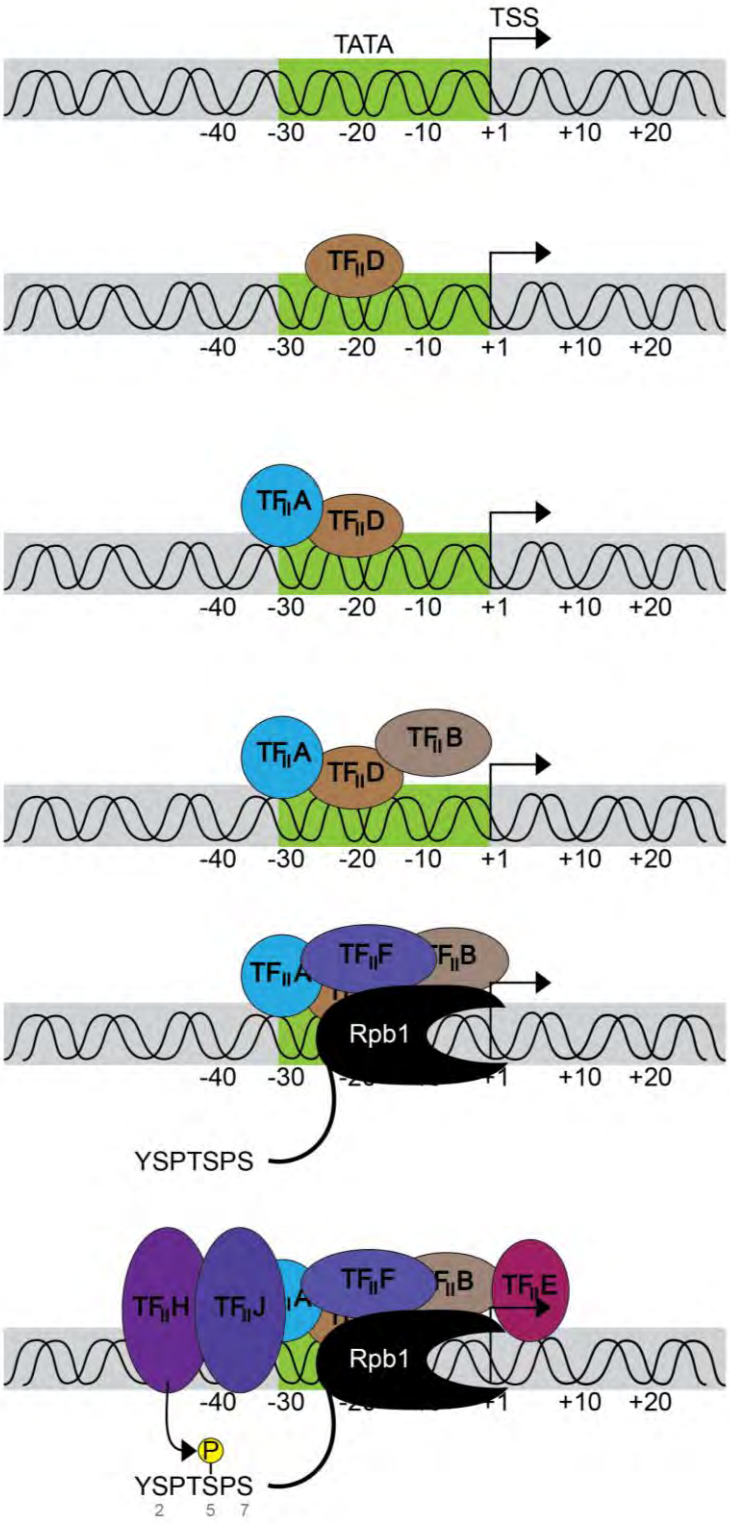

Figure 1.9. Recruitment of the RNA polymerase II to a TATA-box promoter and assembly of pre-initiation complex. The TATA-box is recognised by the TATA binding protein (TBP) that resides within the $TF_{II}D$ complex. $TF_{II}D$ recruits $TF_{II}A$, which in turn recruits $TF_{II}B$. Subsequently, $TF_{II}F$ and $TF_{II}E$ are recruited to the complex. The main polymerase subunit (Rpb1) is then recruited and anchored onto $TF_{II}B$. At this point the initiation complex is assembled and the RNA polymerase produces short abortive transcripts. After the incorporation of factors $TF_{II}H$ and $TF_{II}J$, Rpb1 is phosphorylated at serine 5 of the CTD and the enzyme can proceed into initiation of transcription.

1.2.7.2. Transcription elongation

After the initiation step, the enzyme moves into the gene body. It is however unable to give full length, processed transcripts and pauses around 20bp to 40bp downstream of the TSS (Marshall and Price, 1992). This barrier may be overcome immediately on constitutively expressed genes or it may need a specific signal (Saunders et al., 2006), as is the case for inducible genes. This elongation block is mediated by the action of the negative elongation factor (NELF) and the 5, 6-dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) sensitivity inducing factor (DSIF) (Yamaguchi et al., 1999). In order for the elongation block to be relieved, p-TEFb needs to be recruited to the initiating polymerase, where its cyclin dependent kinase 9 (Cdk9) subunit phosphorylates DSIF and NELF. These phosphorylation events convert DSIF into an activator and cause NELF to be evicted from chromatin (*Figure 1.10*). Finally, p-TEFb phosphorylates the RNA polymerase II CTD at serine 2, a modification that recruits the 5'-capping and the pre-mRNA processing machinery (Bres et al., 2008; Fong and Bentley, 2001). P-TEFb can be recruited by DNA bound co-activators, transcription factors or even chromatin bound activators (reviewed in Peterlin and Price, 2006). Furthermore, the modified CTD provides a tethering point for transcription related chromatin modifiers such as the H3K36 methyltransferase Set2 (Li et al., 2003; Li et al., 2002).

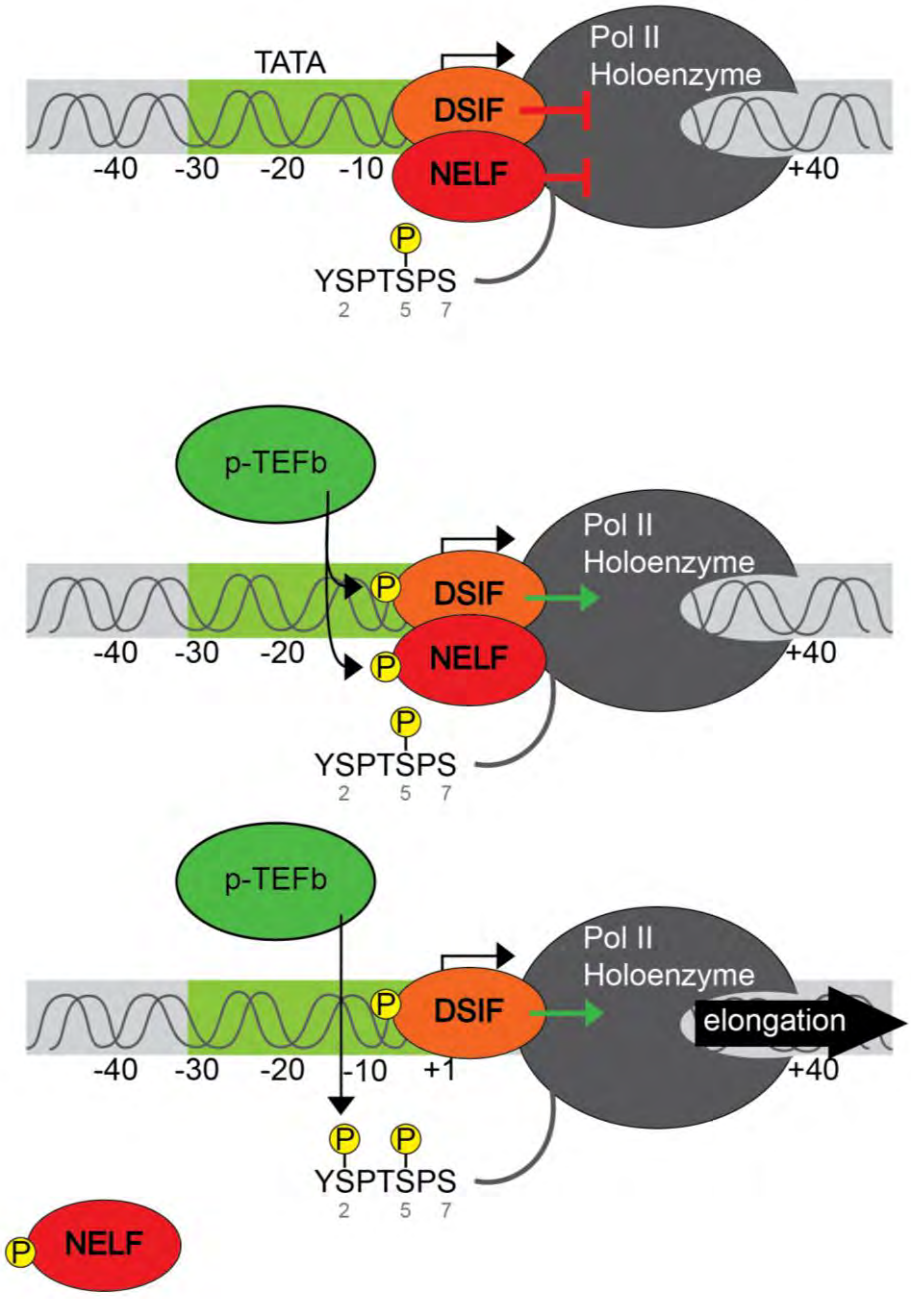

Figure 1.10. RNA polymerase II pause and release into elongation. After the polymerase has initiated transcription, it moves into the gene body and pauses at around +20bp to +40bp. This pause is mediated by two factors, DSIF and NELF. The transcriptional block is alleviated through the signal dependent recruitment of p-TEFb. P-TEFb phosphorylates NELF causing it to be evicted from chromatin. DSIF is also phosphorylated by p-TEFb. In the phosphorylated form and in the absence of NELF, DSIF acts a transcriptional activator. Finally, p-TEFb phosphorylates Rpb1 on the serine 2 residues of the CTD repeats, allowing the enzyme to proceed into productive elongation. The individual TFII factors are not depicted in this schematic as they would make the figure needlessly complicated.

Recruitment of the basal transcription machinery and initiation was until recently considered to be the limiting step in transcription. A growing body of evidence (reviewed in Carrera and Treisman, 2008; Margaritis and Holstege, 2008; Price, 2008) however, suggests that transcription is very often regulated at the transition between initiation and elongation, much like what was initially observed on the Hsp70 promoter in *Drosophila melanogaster (Gilmour and Lis, 1986).*

1.3. DNA methylation

Although the DNA sequence remains mostly unchanged during development, there is another level of information attached to the DNA molecule in the form of DNA methylation. In eukaryotes, methyl-groups are added to cytosine residues and – for animals – predominantly within a CpG dinucleotide. DNA methylation is an ancestral mechanism as it is found in fungi, plants, insects and mammals, albeit with different genomic distributions. This divergence may be indicative of differences in the role DNA methylation plays in different organisms.

The plant *Arabidopsis thaliana* for example has a mosaic pattern of DNA methylation (Deng and Roberts, 2005; Juven-Gershon and Kadonaga, 2010; Lagrange et al., 1998). Repeat elements, heterochromatin, non-transcribing genes and pseudo-genes are all found within DNA-methylated domains. Coding regions of active genes were

also found to be methylated while the promoter and terminator regions of the same genes were non-methylated.

In the fungus *Neurospora crassa*, a mosaic pattern of DNA methylation was also observed (Evans et al., 2001; Kutach and Kadonaga, 2000; Lim et al., 2004). DNA methylation in this case marked repetitive elements and transposons but not the coding regions of transcribed genes.

In vertebrates the majority of CpG dinucleotides in the genome are methylated (Meissner et al., 2008; Singer et al., 1979). This pattern is cell-type specific and is established early during development (reviewed in Borgel et al., 2010; Brandeis et al., 1993)

1.3.1. CpG islands

DNA methylation in animals is limited to cytosines most often in the context of the CpG dinucleotide. Methylated cytosines undergo a spontaneous de-amination reaction, causing the conversion of cytosine into a thymidine residue (Bird, 1980; Coulondre et al., 1978). A proportion of these mutations escapes the DNA repair mechanisms and can be transmitted to the next generation. Over evolutionary time, spontaneous de-amination together with the global CpG methylation pattern observed in vertebrates has resulted in a genome-wide depletion of CpG dinucleotides. In essence the CpG dinucleotide is under-represented in vertebrate genomes compared to the expected frequency.

However, there are regions in the genome where CpG dinucleotides are found at the expected or higher frequency. These regions are on average 1kb long, have GC content of approximately 55% and most often co-localise with gene promoters. These regions were termed CpG islands and are of functional significance for transcriptional regulation. CpG islands can only be observed because they are kept methylation-free, as if they were methylated they would be subject to the same CpG mutational loss as the rest of the vertebrate genome (Bird, 1980; Coulondre et al., 1978). Most importantly, CpG islands are conserved in syntenic regions between humans and mice – that diverged approximately 75 million years ago – suggesting that CpG islands are of functional significance (Deaton and Bird, 2011).

1.3.2. Role of DNA methylation

One of the first things that became evident when DNA methylation was examined in plants and fungi, was that it was targeted to transposable elements through the action of specific protein co-factors (Ohler et al., 2002). This observation gave rise to the “genome defence” hypothesis. DNA methylation was suggested to be a cellular mechanism to prevent activation of transposable elements, that could potentially be detrimental to the genome (Kazazian and Moran, 1998; Kuff and Lueders, 1988; Walsh et al., 1998; Yoder et al., 1997). The fact that mammalian genomes

are globally methylated, with very few exceptions, did not offer any evidence to the contrary (Ehrlich et al., 1982). Moreover, early retroviral transfection experiments, demonstrated that *de novo* methylation of the retroviral DNA in embryonic carcinoma cells could inactivate the virus, while de-methylation could reactivate the virus (Stewart et al., 1982), lending support to the idea that DNA methylation is a defence mechanism against retro-transposons. Indeed, cells that lack DNA methyltransferases exhibit increased transposon expression (Goodrich and Tjian, 2010).

DNA methylation has been employed as a mechanism for transcriptional regulation in mammals. As many as 60% of all mammalian promoters overlap a CpG island (Illingworth et al., 2010; Saxonov et al., 2006). Other CpG islands that do not seem to overlap a protein-coding gene's promoter (the so-called orphan CpG islands) have been in many cases shown to function as promoters for regulatory non-coding RNAs (Panning and Jaenisch, 1996; Sleutels et al., 2002) or act as alternative promoters for known genes (Core et al., 2008; Shiraki et al., 2003). Early reporter-gene based experiments showed that methylation of those CpG island promoters results in transcriptional silencing of the reporter gene, thus establishing a negative correlation between CpG methylation and gene expression (Stein et al., 1982; Vardimon et al., 1982).

This finding posed the question: how does DNA methylation result in transcriptional silencing? The presence of a methyl group in the major groove of the DNA double helix could be excluding binding of sequence specific activator transcription factors, as has been demonstrated for USF

(Watt and Molloy, 1988). This however is not a universal phenomenon, as other factors like SP1 can bind DNA and activate transcription regardless of the methylation status of the underlying DNA sequence (Harrington et al., 1988; Holler et al., 1988).

In addition to that, DNA methylation can exert its repressive function through the recruitment of repressor proteins. Methyl-binding-domain (MBD) protein family members are recruited to the methylated CpG dinucleotides and can recruit repressive complexes that modify the local chromatin environment, thus making it unfavourable for transcription (Fuks et al., 2003b; Jones et al., 1998; Nan et al., 1998; Prokhortchouk et al., 2001; Yoon et al., 2003a). The common theme emerging from a vast number of studies by several research groups is that methylation of a CpG island promoter results in transcriptional silencing of the associated gene. However, the non-methylated state of the CpG island promoter does not necessarily result in active transcription (Bird et al., 1987; Weber et al., 2007). In fact, numerous genes that are regulated by CpG island promoters become transcriptionally silenced during mammalian development. This transcriptional silencing is only rarely accompanied by DNA methylation (reviewed in Suzuki and Bird, 2008).

Contrary to most genes, imprinted genes and genes that reside on the inactive X chromosome in mammals gain DNA methylation on their CpG island promoters. DNA methylation of those promoters is crucial for the maintenance of transcriptional silencing (Chen and Li, 2004; Dodge et al., 2005; Karpf and Matsui, 2005).

DNA methylation is also involved in the establishment of lineage appropriate expression profiles during differentiation (Deaton et al., 2011; Farthing et al., 2008; Illingworth et al., 2008; Mohn et al., 2008; Schilling and Rehli, 2007; Weber et al., 2007; Yagi et al., 2008). These studies clearly demonstrated that DNA methylation patterns change during cell differentiation from embryonic stem cells to terminally differentiated cells.

This conversion of DNA methylation patterns is dependent on DNA methyltransferases (DNMTs) and proteins of the TET family (Ito et al., 2010; Wu et al., 2011). TET proteins can hydroxylate methyl-cytosine residues and convert them to hydroxymethyl-cytosine (Ito et al., 2011; Tahiliani et al., 2009). Hydroxymethyl-cytosine can then be removed from the genome and replaced by unmodified cytosine. This process is dependent on the base excision repair mechanism (He et al., 2011; Maiti and Drohat, 2011). Additionally, cytosine hydroxymethylation itself may negate the repressive effects of cytosine methylation, as the methyl-binding proteins of the MBD family are unable to bind hydroxymethyl-cytosine and exert their repressive functions (Jin et al., 2010).

1.3.3. Chromatin signature of a CpG island

CpG islands differ significantly from the surrounding DNA. Firstly, they can be separated from the rest of the genome solely by their nucleotide composition. These regions present a GC percentage of approximately 55% as compared to 42% in bulk mouse genome (41% in the human genome). Moreover, the CpG dinucleotide is vastly overrepresented in comparison to the CpG depleted mammalian genomes.

These primary sequence characteristics can already influence the assembly of those sequences into nucleosomes. It was observed that certain dinucleotide combinations (especially ApA and TpT) are found in the genome with approximately a 10bp periodicity, the same as the helical periodicity of the DNA molecule. This was proposed to aid the bending of the DNA double helix around the core nucleosome particle (Trifonov and Sussman, 1980). As the physical properties of individual nucleotides differ, it is conceivable that different dinucleotide combinations could introduce a kink in the DNA double helix, thus aiding the wrapping of DNA around the histone core. A/T pairs are underrepresented in CpG islands, thus rendering DNA locally rigid and less able to stably assemble into nucleosomes. Indeed, more recent studies have demonstrated a general nucleosome depletion over CpG islands (Heintzman et al., 2007; Ozsolak et al., 2007; Schones et al., 2008).

The differences of CpG islands to the rest of the genome are not limited to the nucleotide composition and nucleosome depletion. The nucleosomes that are present on or adjacent to a CpG island are marked by trimethylation of lysine 4 on the N-terminal tail of histone H3 ($H3K4me_3$). This mark is thought to be a mark of active transcription (reviewed in Kouzarides, 2007). The deposition of this mark is thought to be mediated by CFP1, a CxxC domain protein (Thomson et al., 2010). The CxxC domain of CFP1 can bind to non-methylated CpG dinucleotides (Voo et al., 2000) and CFP1 has been shown to interact with several H3K4 specific methyltransferases (Ansari et al., 2008; Lee and Skalnik, 2002, 2005). The presence of the $H3K4me_3$ mark does not depend on the promoter activity of the CpG island, as artificial CpG islands with no promoter activity readily gained the $H3K4me_3$ mark when inserted in murine ES cells (Thomson et al., 2010).

This distinct chromatin structure of CpG island promoters makes them favourable for transcription by recruitment of additional histone modifying complexes to the $H3K4me_3$ mark and by possibly negating the need for nucleosome remodelling.

1.4. Embryonic stem cells

1.4.1. Origins of the embryonic stem cell

When a fertilized egg starts dividing it gives rise to cells that are completely identical in every aspect (*Figure 1.11*). After the 8 cell stage, during blastocyst formation, the first steps of differentiation are becoming evident, as some the cells of the conceptus form the trophectoderm – a tissue that will not be part of the fetus – and the inner cell mass (ICM). The inner cell mass is the part of the conceptus that will develop into the fetus. ICM cells are pluripotent, meaning they are able to generate all tissues found in the adult organism but cannot generate extraembryonic tissue. Such cells can be isolated from the developing embryo and they can be cultured *in vitro* (Evans and Kaufman, 1981). Once isolated they are called embryonic stem cells (ES cells) which – under the appropriate culture conditions – exhibit unlimited self-renewal capacity and extensive differentiation potential, reflecting their origins.

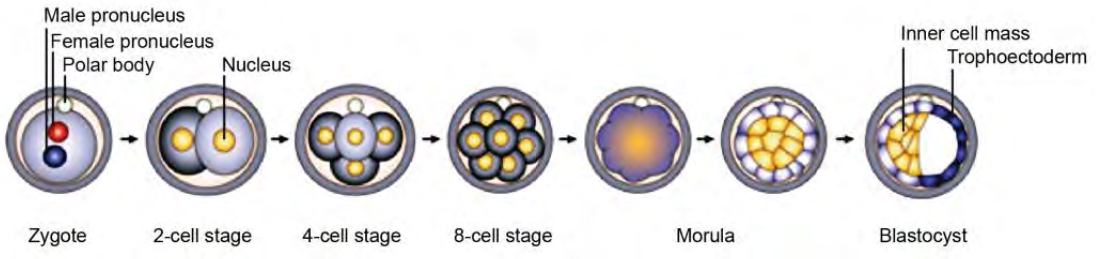

Figure 1.11. Development of the zygote to the blastocyst (Figure adopted from Zernicka-Goetz, 2005).

1.4.2. Pluripotency and lineage commitment

As the embryo develops further, the individual cells in the ICM activate different transcriptional programs depending on cues from neighbouring cells. In essence, different genes are transcribed into messenger RNA (mRNA), therefore different proteins are synthesised in the different cells. These differences in gene expression underlie all differentiation processes (Mansergh et al., 2009). All the ICM cells start off as pluripotent, but as they progress through development their differentiation potential is gradually limited until finally a specific mature cell type is formed (reviewed in Reik, 2007). Techniques have been developed that recapitulate the differentiation of the ICM cells by culturing ES cells *in vitro* under specific culture conditions that will drive differentiation toward a specific cell type (Doetschman et al., 1985; Fehling et al., 2003; Lancrin et al., 2009; Pearson et al., 2008; Wobus et al., 1988).

1.4.3. Bivalent chromatin domains and the poised polymerase

The chromatin fibre within a cell exists in several different states that impact greatly on its ability to be transcribed into mRNA. The presence of activating or repressive transcription factors, post-translational modifications on the histone molecules and the resulting degree of compaction of the chromatin fibre are the key factors that impact on

whether a gene can be transcribed or not (reviewed in Bannister and Kouzarides, 2011; Lee and Young, 2000).

Genes that are not required usually reside in compacted chromatin domains in mature cell types (Mikkelsen et al., 2007). The situation in ICM/ES cells is quite different. In ICM and ES cells the vast majority of the genome is accessible for the transcriptional machinery (Zeitlinger et al., 2007). The pluripotency of these cells is reflected on the molecular level, where specification genes of different lineages are not expressed but harbour both activating and repressive chromatin modifications (Bernstein et al., 2006). Moreover, the transcriptional machinery is in place and waiting for the environmental cue to transcribe the genes in question (Zeitlinger et al., 2007). These domains have been termed bivalent chromatin domains (Bernstein et al., 2006).

More recent data suggest that bivalent domains are not as prevalent as initially thought and part of the observed bivalency may be a result of culture conditions (Marks et al., 2012). Depending on the cue received, a subset of those genes will resolve in the active state driving the cell in a specific differentiation path, while all the inappropriate genes will be silenced and eventually compacted (Mikkelsen et al., 2007).

Put simply, ES cells keep all their options open until they receive an environmental signal. At that point they undertake a cell fate decision and start progressing through differentiation – via a series of subsequent binary decisions – by adjusting their transcriptional program under the influence of newly expressed lineage specific transcription factors.

1.4.4. ES cells as a model system for biological processes

ES cells are a good model to study developmental processes *in vitro* as they retain the differentiation potential of the ICM cells. However, the implications of the establishment of ES cell lines extend beyond that. ES cells can be grown in vast numbers in the lab due to their very high proliferation rates and enhanced self-renewal capacity, without the need for immortalization. ES cells can, relatively easily, be genetically modified to generate knock-out/in ES cells, tag proteins to simplify further analyses (Joyner et al., 1989; Smithies et al., 1985; Thomas and Capecchi, 1987; Thomas et al., 1992) and then either used as a cell line or micro-injected into a developing embryo, to yield chimeric embryos (Bradley et al., 1984). These chimeric animals can then be examined for germline transmission of the modified genes and used to establish pure genetically modified strains. This technology is now quite advanced in the mouse (Hofemeister et al., 2011) and has generated a large number of different mouse lines harbouring mutations, constitutive or inducible deletions and various tagged proteins (reviewed in Skarnes et al., 2011).

1.5. Histone modifications in development

Histone modifications have proven to be a key aspect of transcriptional regulation (as exemplified in paragraph 1.2). Although the drivers of differentiation are transcription factors, the enzymatic activities depositing such modifications also have an impact on differentiation processes. Several studies have reported severe developmental defects and/or embryonic lethality as a result of mutations or deletion of (among others) the histone acetyltransferases MOF (Gupta et al., 2008), p300 (Yao et al., 1998), GCN5 (Xu et al., 2000; Yamauchi et al., 2000) and the histone methyltransferases MLL1 (Yu et al., 1995) and MLL2 (Glaser et al., 2006).

1.5.1. The polycomb / trithorax antagonism

In *Drosophila melanogaster*, mutation of the *Polycomb* gene gives rise to a homeotic phenotype (Lewis, 1978). *Polycomb* is not related to the homeobox genes that control embryo pattering but was shown to be a repressor of the bithorax gene cluster that controls formation of the posterior part of the fly (Lewis, 1978, 1982; Wedeen et al., 1986). A group of genes (Polycomb Group, PcG) was discovered later that when mutated presented similar homeotic phenotypes (Jurgens, 1985).

In the following decades, members of the PcG were shown to encode for proteins that form 3 distinct complexes, the polycomb repressive complex 1 (PRC1), polycomb repressive complex 2 (PRC2) and the pleiohomeotic repressive complex (PhoRC), with distinct histone modifying activities (presented briefly in paragraph 1.2.6.3 and reviewed in Margueron and Reinberg, 2011; Schwartz and Pirrotta, 2007). Additionally, these complexes have been shown to act on a much wider scale than just the homeobox gene clusters (reviewed in Schuettengruber et al., 2007; Schwartz and Pirrotta, 2007).

The gene *Trithorax* was shown to have a suppressive effect on PcG mutant phenotypes (Ingham, 1983). Later, a large number of genes with similar activity were characterised and classified as the Trithorax group (TrxG) (reviewed in Schuettengruber et al., 2007; Schuettengruber et al., 2011). TrxG proteins generally oppose the action of PcG proteins and they – like the PcG proteins – are not limited to homeobox gene clusters (reviewed in Schuettengruber et al., 2007). TrxG proteins however are not as biochemically homogeneous as the PcG. They contribute to many different transcriptional activating complexes with roles in histone modification and chromatin remodelling (reviewed in Schuettengruber et al., 2011). TrxG proteins have been implicated in ES cell self-renewal (Ang et al., 2011), cell fate decisions and proliferation (Bagchi et al., 2007), programmed cell death (Tyagi and Herr, 2009) and X-chromosome inactivation (Pullirsch et al., 2010).

1.5.2. The SET1 / MLL family

A subset of the TrxG is the SET1/MLL protein family. These proteins have been shown to possess histone methyltransferase activity targeted to lysine 4 of histone H3 (Goo et al., 2003; Hughes et al., 2004; Milne et al., 2002; Wysocka et al., 2003; Yokoyama et al., 2004). These findings offered a molecular explanation to the PcG – TrxG antagonism that has been observed genetically. This antagonism is manifested by the opposing functional outcomes of the repressive methylation of lysine 27 on histone H3 (Cao et al., 2002) and ubiquitination of lysine 121 on histone H2A (Wang et al., 2004a) mediated by PcG proteins and the activating methylation of lysine 4 on the same histone catalysed by members of the SET1/MLL family.

The SET1/MLL family is comprised of at least six different proteins in mammals. Those are MLL1 (KMT2A, HRX), MLL2 (KMT2D, Wbp7), MLL3 (KMT2C, HALR), MLL4 (KMT2B, ALR), SET1A (KMT2E) and SET1B (KMT2F). The different SET1/MLL members are illustrated in *Figure 1.12.* The nomenclature for MLL proteins is rather confusing as mouse MLL2 (Wbp7) was named MLL4 in humans and mouse MLL4 (ALR) was named MLL2 in humans. In this study we are exclusively using the mouse nomenclature. *In silico* phylogenetic analysis revealed that these proteins exist as orthologous pairs. The MLL1-2 pair is the closest homologue of the *D. melanogaster Trithorax* (FitzGerald and Diaz, 1999; Glaser et al., 2006).

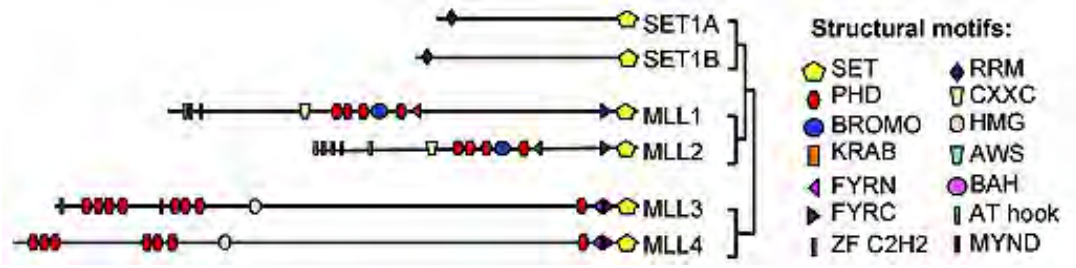

Figure 1.12. The SET1/MLL protein family (Figure adapted from Ruthenburg et al., 2007). The SET1/MLL family of proteins is composed of three orthologous pairs of proteins.

Set1 – the only histone H3 lysine 4 methyltransferase in yeast – is recruited to the initiating RNA polymerase II via the Paf1 complex (Ng et al., 2003). The human PAF1 complex can also recruit histone H3 lysine 4 methyltransferase activities to the RNA polymerase (Rozenblatt-Rosen et al., 2005), but a later study (Milne et al., 2010) demonstrated that, at least *in vitro*, this interaction is specific to MLL1.

MLL3 and MLL4 appear to be specifically recruited to target promoters by transcription factors and the ASC2 co-activator (Demers et al., 2007; Lee et al., 2006; Mo et al., 2006; Patel et al., 2007). Deregulation of MLL3 or MLL4 has been implicated in solid tumour development (Lee et al., 2009; Vakoc et al., 2009; Wang et al., 2011). MLL3 has been reported to be required for normal adipogenesis (Lee et al., 2008). MLL4 has also been implicated in leukaemias (reviewed in Damm et al., 2012), while MLL4 mutations have been attributed a direct causal role in Kabuki syndrome (Ng et al., 2010).

MLL1 is the most studied MLL protein, mainly due to the fact that it causes aggressive leukaemias when fused to a variety of other proteins (reviewed in Collins and Rabbitts, 2002; Daser and Rabbitts, 2005; and originally described in Djabali et al., 1992; Tkachuk et al., 1992). The leukaemogenic potential of the MLL1 fusion proteins is mediated by aberrant recruitment of elongation promoting factors (reviewed in Slany, 2005, 2009; Smith et al., 2011). This elongation complex recruits p-TEFb resulting in phosphorylation of the RNA polymerase II CTD and loss of pausing control. Further studies demonstrated that the non-methylated CpG binding capacity (mediated by the CxxC domain) of MLL1 is maintained in the MLL1 fusion proteins and is required for leukaemogenic transformation (Ayton et al., 2004). Consistent with the haemopoietic deregulation observed in the presence of aberrant forms of MLL1, wild type MLL1 is required for normal haemopoiesis. Several studies have demonstrated that MLL1 is absolutely required for both primitive and definitive haemopoiesis and specifically for the generation of haemopoietic stem cells (Ernst et al., 2004; Hess et al., 1997; Jude et al., 2007; McMahon et al., 2007; Yagi et al., 1998).

MLL1 has been implicated in transcriptional memory (Blobel et al., 2009). This study presents evidence that MLL1 is redistributed to highly transcribed genes before mitosis and remains associated with the mitotic chromosomes. This redistribution facilitates robust reactivation of the genes after cytokinesis.

Biochemical purification of the MLL1 complex identified the product of the *MEN1* gene as a part of the MLL1 complex (Hughes et al., 2004; Yokoyama et al., 2004). Yokoyama et al. (2004) demonstrated that *MEN1* deletion phenocopies the effects of *Mll1* deletion suggesting a key role of the *MEN1* gene product in the functional integrity of the MLL1 complex. The other subunits of the core MLL1 complex are WDR5, RBBP5 and ASH2L and are shared between the MLL1 and MLL2 complexes (Dou et al., 2006; Wysocka et al., 2003). These studies demonstrated that loss of any of these subunit results in partial or total loss of MLL1 complex methyltransferase activity both *in vitro* and *in vivo*. Additionally, the MLL1 complex has been shown to interact with the histone H4 lysine 16 acetyltransferase MOF (Dou et al., 2005), the histone acetyltransferase CBP (Ernst et al., 2001) and members of the SWI/SNF family of chromatin remodelers (Rozenblatt-Rosen et al., 1998). MLL1 appears to be part of a mega-complex that concentrates several activities that can co-operatively activate transcription. Milne and colleagues (Milne et al., 2005b) provided further support to this idea by demonstrating that deletion of *Mll1* resulted in abnormal RNA polymerase II distribution and changes in histone modification levels on a subset of the homeobox gene promoters. Disruption of either the MLL1 or MLL2 complex resulted in depletion of the histone H3 lysine 4 trimethyl mark on specific homeobox gene promoters and concomitant loss of the basal transcription machinery in mouse embryonic fibroblasts (Wang et al., 2009). Another study (Terranova et al., 2006) demonstrated that mouse embryos lacking MLL1 exhibit abnormal

homeobox gene silencing, along with aberrant DNA methylation of those genes' promoters. Studies published by the Timmers laboratory (van Ingen et al., 2008; Vermeulen et al., 2007) demonstrate that the basal transcription factor $TF_{II}D$ can bind directly to trimethylated lysine 4 of histone H3. $TF_{II}D$ is a key basal transcription factor that can potentialy nucleate the rest of the basal transcriptional machinery and trigger the assembly of the RNA polymerase II pre-initiation complex (reviewed in Buratowski, 1994). If $TF_{II}D$ recruitment and/or stabilisation depended on the presence of the $H3K4me_3$ mark, loss of this mark would result in transcriptional down-regulation or complete silencing of genes that operate through this mechanism.

Taken together, these results suggest that MLL proteins may be recruited to promoters prior to the basal transcription machinery (unlike yeast Set1) and function by i) setting up a permissive chromatin landscape for transcription, ii) recruiting the basal transcription machinery and iii) preventing DNA methylation. Direct evidence for this idea however has been so far lacking.

1.5.3. MLL2 structure and function

MLL2 is a 2713 amino-acid protein with a very similar domain composition to MLL1 (*Figures 1.12 and 1.13*). The functions attributed here to the MLL2 domains are based on their similarity to other proteins, as there is no direct experimental evidence on their function in the context of MLL2. MLL2 harbours 3 AT hook domains that bind the minor groove of the DNA double helix in AT-rich regions (Reeves and Nissen, 1990), 2 speckled nuclear localization signals that dictate the subnuclear localization of the protein (Hedley et al., 1995), a CxxC type zinc finger that binds to non-methylated CpG dinucleotides (Ayton et al., 2004; Birke et al., 2002; Lee et al., 2001), 3 PHD fingers that can bind histone H3 trimethylated at lysine 4 (Chang et al., 2010), an atypical bromodomain that would normally bind acetylated histones (Dhalluin et al., 1999; Jacobson et al., 2000; Owen et al., 2000), an extended PHD finger and a SET domain that catalyses trimethylation of lysine 4 on histone H3 (Dou et al., 2005; Milne et al., 2002; Nagy et al., 2002; Roguev et al., 2001). MLL2 also contains a taspase 1 cleavage site and 2 FY-rich regions. In the case of MLL1, the FY regions have been shown to mediate re-association of the two MLL1 parts after taspase-mediated cleavage (Hsieh et al., 2003).

Unlike MLL1, MLL2 function is rather poorly characterised. Knock-out experiments have demonstrated that MLL2 is indispensable for mouse embryonic development, as embryos lacking MLL2 die before embryonic day 11.5 (Glaser et al., 2006). Notably, the *Mll2* knock-out embryos die

due to widespread apoptosis and general tissue disorganisation rather than specfic developmental defects. ES cells lacking MLL2 are viable, albeit with defects in differentiation and slightly increased apoptosis rates (Lubitz et al., 2007). These observations reveal a crucial role for MLL2 in early embryonic development.

Using a conditional knock-out allele for *MLL2*, a more recent study has demonstrated that MLL2 is only required briefly during development. (Glaser et al., 2009). Postnatal deletion of *Mll2* did not have any severe effects at the organism level. However, gametes were affected by loss of MLL2, which is required for maturation of spermatocytes (Glaser et al., 2009) and bulk H3K4 trimethylation and survival in oocytes (Andreu-Vieyra et al., 2010). Additionally, it has been demonstrated recently that MLL2 plays a role in macrophage function (Austenaa et al., 2012). According to Austenaa et al. MLL2 is absolutely required for transcription of the *Pigp* gene, which encodes for a critical enzyme in the GPI anchor glycolipid synthesis. As a result *Mll2* knock-out macrophages exhibit a complete lack of GPI-anchored membrane proteins and impaired lipo-polysaccharide responses.

Glaser et al. (2009) also demonstrated that only a small number of genes are significantly down-regulated in response to *Mll2* deletion, most prominent of which is *Magoh2* – an until then uncharacterised gene. *Magoh* and *Magoh2* (also known as *MagohB*) code for identical proteins except for a single aminoacid substitution (prediction from the *in silico* translated mRNA sequence, ncbi BLASTx). The *Drosophila* homologue of

Magoh and *Magoh2* is *mago nashi*. The *mago nashi* gene product is a component of the exon-exon junction complex (Kataoka et al., 2001). *Magoh2* expression is regulated by a CpG island promoter and it is expressed in all the tissues examined so far, including ES cells (Glaser et al., 2009). The same study demonstrated that *Magoh2* expression is completely abolished upon *Mll2* deletion, while *Magoh* expression was unaffected. *Magoh2* is exquisitely dependent on MLL2, with no functional redundancy from other SET1/MLL family members. *Magoh2* transcriptional silencing is accompanied by depletion of histone H3 lysine 4 trimethylation and an increase in DNA methylation over the CpG island promoter.

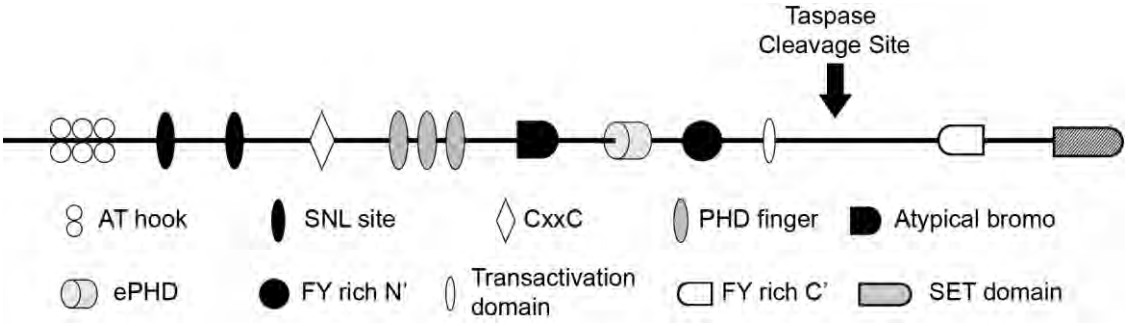

Figure 1.13. Structure of the MLL2 protein (Figure re-drawn from Glaser et al., 2009). MLL2 is cleaved by Taspase 1 and the 2 parts hetero-dimerise via the two FY-rich regions. MLL2 contains 3 AT hooks, 2 speckled nuclear localisation sequences (SNL), a CxxC zinc finger, 3 PHD fingers, an atypical bromo-domain, an extended PHD finger (ePHD) and the catalytic SET domain.

1.6. Aims of the present study

Currently, besides biochemical information, very little is known about the molecular mechanism by which individual MLL complexes regulate their target genes in living cells. The reason for this is that different SET1/MLL complexes can compensate for the lack of individual members. This was recently confirmed by chromatin immunoprecipitation coupled with high throughput sequencing experiments demonstrating that more than one SET1/MLL family members can be recruited to the same gene (*unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). The same study also showed that the majority of MLL2 direct targets are genes regulated by CpG island promoters. However, the role of MLL complexes in controlling the activity of such promoters is poorly understood. Studies described above have demonstrated that MLL1 depletion leads to transcriptional silencing accompanied by loss of histone H3 lysine 4 methylation, histone acetylation and RNA polymerase II. This silencing is often stabilised by DNA methylation. Although Glaser et al. (2009) provided some evidence suggesting that MLL2 operates by a similar mechanism, there is no data on whether *Mll2* deletion leads to the same chromatin state switch and RNA polymerase II eviction. To this end, this study examines the fine chromatin structure of the *Magoh2* promoter in *Mll2*$^{-/-}$ murine ES cells.

Glaser et al. (2009) reported that *Magoh2* transcriptional silencing is accompanied by DNA methylation. It is unclear however, whether DNA methylation precedes transcriptional silencing or merely maintains it. Moreover, it is not known where the basal transcription machinery is situated within this hierarchy. To gain insight into these mechanistic details, this study employs time-course experiments on an inducible *Mll2* knock-out system to try and establish the order of events that lead to *Magoh2* silencing and DNA methylation. Additionally, there is no information on whether this epigenetic silencing can be reversed. This issue is addressed herein by re-introduction of wild type or mutant MLL2 and examination of the *Magoh2* transcriptional status.

Many studies have reported that MLL1 deletion results in dramatic haemopoietic deficiencies. Our knowledge regarding MLL2 in haemopoiesis is limited to the fact that *Mll2* is required for macrophage responses (Austenaa et al., 2012). MLL2 could play a more fundamental role in early haemopoiesis that is masked by the widespread apoptosis and early lethality observed in $Mll2^{-/-}$ embryos. This issue is addressed by *in vitro* differentiation of ES cells toward haemopoietic progenitors or monocytes, with deletion of *Mll2* at different stages of the haemopoietic differentiation process.

2. MATERIALS AND METHODS

2.1. Cell lines

E14 wild type, *Mll2*$^{-/-}$, *Mll2*$^{F/F}$, *Mll2*$^{F/+}$ (Glaser et al. 2006, 2009; Lubitz et al. 2007) and mutant MLL2 transfected *Mll2*$^{F/F}$ ES cells (generated by Mr. A. Gupta and Mr. D.C. Torres, not published) were provided by Dr. Stewart's lab, Technische Universitaet Dresden. *Sp1*$^{-/-}$, *Sp1R* and *Sp3*$^{-/-}$ ES cells were provided by Dr. Philipsen, Erasmus University Rotterdam (Akalin et al., 2009; Lenhard et al., 2012).

2.2. Cell Culture

2.2.1. Routine murine embryonic stem cell culture

Murine embryonic stem (ES) cells were grown in high glucose DMEM (Sigma) supplemented with 15% batch tested serum for maintenance of pluripotency (PAA Gold FBS), 1000 U/ml leukaemia inhibitory factor (LIF) (ESGRO, Millipore), 25 mM HEPES (Gibco), 1 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco), 100 U/ml penicillin / 100 µg/ml streptomycin (Gibco), 1× non-essential amino-acids (Sigma), 0.15 mM mono-thioglycerol (MTG, Sigma) at 37^{o}C, 5% CO_2, on tissue culture grade plastics (Falcon, Corning) at a plating density of

10^4 cells/cm^2. Cells were split every 2-3 days, by removing the medium, washing very briefly with phosphate buffered saline (PBS, Invitrogen) and brief trypsinization in an appropriate volume of 1× trypsin / ethylene diamine tetra-acetic acid (EDTA) (Gibco) at room temperature. Trypsin was inactivated by addition of an equal volume of serum-containing medium. Cells were counted in an improved Neubauer haemocytometer (Hawksley), under an Olympus BX45 microscope.

2.2.2. Differentiation of ES cells in semi-solid media

ES cells were harvested and washed in PBS to remove any remaining LIF. Subsequently, they were resuspended in 1.2% base methylcellulose (Stemcell Technologies) in IMDM medium (Invitrogen) containing 10% serum batch tested for embryoid body differentiation (PAA FBS), 10% M-CSF conditioned medium, 5% IL3 conditioned medium, 0.15 mM MTG, 2 mM L-glutamine and 100 U/ml penicillin / 100 µg/ml streptomycin, at 10^3-10^5 cells/ml. 4ml were plated in 60 mm low adherence bacteriological grade plates (Sterilin) and grown at 37^{o}C / 5% CO_2 for 15-20 days. Embryoid bodies (EB) were checked regularly for blood island formation and monocyte production on an Olympus CKX41 inverted microscope, fitted with an Olympus C7070 digital camera.

2.2.3. Differentiation of ES cells to blood progenitors via haemogenic endothelium

ES cells were harvested by trypsinization from the routine culture medium and plated on gelatinized (0.1% pork skin gelatine (Sigma) in PBS for 20 minutes at room temperature) tissue culture grade containers at a density of 2-4×10^4 cells/cm^2 in standard ES growth medium (described above). 24 hours later the cells were trypsinised again and plated on gelatinized containers at a density of 2-4×10^4 cells/cm^2 in IMDM (Sigma) supplemented with 15% ES serum, 1000 U/ml LIF, 25 mM HEPES, 1 mM sodium pyruvate, 2 mM L-glutamine, 100 U/ml penicillin / 100 µg/ml streptomycin, 1× non-essential amino-acids, 0.15 mM MTG. 24 hours later the cells were harvested by brief trypsinization and washed once with PBS to remove all remaining ES medium and LIF. The cells were resuspended in IMDM based *in vitro* differentiation (IVD) medium containing 15% serum batch tested for EB differentiation (PAA FBS), 100 U/ml penicillin / 100 µg/ml streptomycin, 0.15 mM MTG, 180 µg/ml human transferrin (Roche) and 50 µg/ml L-ascorbic acid (Sigma) at a density of 5×10^4 cells/ml. The cells were plated in low adherence bacteriological dishes (Corning) and incubated for 3.5 days at 37^{o}C / 5% CO_2 to form EBs. The EB suspension was harvested, transferred to appropriate tubes and the EBs were left to settle by gravity for 10 minutes. The medium was discarded, the EBs were washed with PBS and left to settle again by gravity. PBS was removed and the EBs were completely dissociated by

addition of 2 ml 1× Trypsin/EDTA and gentle pipetting. Trypsin was inactivated by adding an equal volume of IMDM containing 20% EB serum. The cells were counted and centrifuged at 300×g for 5 minutes at room temperature and then resuspended in 200 µl MACS buffer: 0.5% bovine serum albumin (BSA, Miltenyi), 2 mM ethylene-diamine-tetraacetic acid di-sodium salt (EDTA, Sigma) in PBS. A small amount of cells was kept separately and stained with 0.5 µg of a phycoerythrin (PE) conjugated α-FLK1 antibody (eBioscience) or an isotype control antibody (eBioscience) in 100 µl MACS buffer to verify FLK1 expression by fluorescence activated cell sorting (FACS) on a BD Bioscience LSRII instrument. FACS data were analyzed using the BD FACS Diva software suite. The remaining cells were then stained for FLK1 with 1 µg of a biotin conjugated α-FLK1 antibody (eBioscience) for 15 minutes on ice. The cells were washed twice with MACS Buffer and then bound to anti-biotin magnetic beads (Miltenyi) according to the manufacturer's instructions. The FLK1 positive haemangioblast-enriched cells were purified using a Miltenyi AutoMACS instrument. The cells in the positive and negative fractions were counted and compared to the percentages acquired from FACS analysis. Purified FLK1 positive cells were centrifuged at 300×g for 5 minutes at room temperature and resuspended in an appropriate volume of IMDM based blast medium containing 10% EB FCS, 2 mM L-glutamine, 100 U/ml penicillin / 100 µg/ml streptomycin, 0.15 mM MTG, 180 µg/ml human transferrin, 50 µg/ml L-ascorbic acid, 20% D4T conditioned medium, 10 µg/ml VEGF (Peprotech) and 10 µg/ml IL-6 (Peprotech) and

plated in tissue culture grade containers at a density of 1-2×10^4 cells/cm^2 and incubated for a further 3 to 4 days to allow for haemogenic endothelium formation. The whole blast culture was harvested at the timepoints indicated in the main text by brief trypsinization and the cells were resuspended in PBS. A fraction of the cells was separated from the bulk (that was used for RNA isolation), distributed into 6 FACS tubes and centrifuged at 300×g for 5 minutes at room temperature. The cells were resuspended in 100 µl MACS buffer. The 6 samples prepared were either 1) left unstained, or stained with 2) 1 µl PE-Cy7–α-CD41 antibody, 3) 0.5 µl PE–α-Tie2 antibody, 4) 0.5 µl PE–IgG isotype control antibody, 5) 1 µl PE-Cy7– / 1 µl APC–IgG isotype control antibodies and 6) 1 µl PE-Cy7–α-CD41 / 0.5 µl PE–α-TIE2 / 1 µl APC–α-C-KIT antibodies. All the antibodies used were from eBioscience. Staining was performed in the dark for 20 minutes on ice. Subsequently the cells were washed in 1 ml MACS buffer, resuspended in 500 µl MACS buffer and analyzed on a LSRII instrument. Post-acquisition data analysis was performed using the BD FACS Diva software suite.

2.2.4. Genotyping

The *Mll2*$^{F/F}$ and *Mll2*$^{F/+}$ ES cells were tested for recombination of the *Mll2* alleles twice by Southern blot (see main text) and after that by PCR every time they were induced with 4-hydroxytamoxifen (OHT). OHT was added to the culture medium at a final concentration of 10^{-7} M. The genotyping PCR was performed using the primers shown in *table 2.1.* The 145se/147as primer pair detects all *Mll2* alleles while the 145se/146as primer pair detects only the larger targeted (F) and WT alleles. This pair was used to verify deletion of the *F* allele as PCR with the 145se/147as pair was strongly biased for the smaller *Fc* product.

Table 2.1. Primers used to confirm deletion of the *Mll2* F allele.

Primer	Sequence	Tα°C	Amplicon
145se	CGGAGGAAGAGAGCAGTGACG	65	F: 1500bp, WT: 1400bp, Fc: 800bp
147as	GGACAGGAGTCACATCTGCTAGG		
145se	CGGAGGAAGAGAGCAGTGACG	65	F: 1132bp, WT: 1032bp
146as	GGGACCGAAGCGCAGAGC		

2.2.5. Disruption of RNA polymerase II by α-amanitin and DRB

DRB blocks transcriptional elongation by inhibiting p-TEFb enzymatic activity (Fraser et al., 1978; Marshall et al., 1996). α-amanitin binds the active site of RNA polymerase II and this complex is subsequently targeted for degradation (Nguyen et al., 1996). Mouse ES cells were grown under standard conditions. 5,6-dichlorobenzimidazole 1-β-D-ribofuranoside (DRB, Sigma) or α-amanitin (Sigma) were added to the culture medium to give a final concentration of 100 μM or 5 μg/ml, respectively. The cells were grown for up to 24 hours in DRB and up to 48 hours in α-amanitin and then harvested using a method suited for the downstream analyses.

2.3. Apoptosis assays by annexin V staining

Mll2$^{F/F}$ and *Mll2*$^{F/+}$ ES cells were grown under standard conditions. OHT was added to the culture medium from a stock of 10^{-3} M in 100% ethanol to yield a final concentration of 10^{-7} M. Control cells (labelled -OHT in main text) were treated with an equal volume of 100% ethanol. OHT was added on subsequent days so that all the timepoints could be harvested at once. The cells were stained with FITC conjugated Annexin V and propidium iodide (PI) using an Annexin V apoptosis detection kit from Santa Cruz Biotechnology according to the manufacturer's instructions. The stained

cells were analysed by FACS on a BD Bioscience LSRII instrument. FACS data were analysed using the BD FACS Diva software suite.

2.4. Cell cycle assays by propidium iodide staining of fixed cells

Mll2$^{F/F}$ and *Mll2*$^{F/+}$ ES cells were grown under standard conditions and induced with OHT as described above. OHT (or ethanol where appropriate) was added to all the cells on the same day and the individual time-points were harvested sequentially. The cells were harvested by brief trypsinization and counted. 10^7 cells were transferred to 1.5 ml tube and centrifuged at 300×g for 5 minutes at 4^{o}C. The cells washed once with ice cold PBS, centrifuged again and resuspended in the residual volume (30-50 µl) of PBS. The cells were fixed by addition of 1 ml 70% methanol that was pre-cooled to -20^{o}C. The fixed cells were kept at -20^{o}C until the complete set of time-points was collected. The cells were centrifuged at 400×g for 5 minutes at room temperature, washed once with PBS and resuspended in 500 µl of freshly prepared PBS containing 50 µg/ml propidium iodide and 10 µg/ml RNase A. The cells were stained for 30 minutes at room temperature in the dark and analyzed on a DB Bioscience LSRII FACS instrument. The post-acquisition data analysis was performed using the Verity ModFit LT 3.2 software suite.

2.5. Transfections

All transfections were performed in 6-well plates (Falcon) using Lipofectamine LTX with PLUS reagent from Invitrogen according to the manufacturer's instructions. Antibiotic selection was optimized by determining the minimum concentration sufficient to kill non-transfected cells. The concentrations that were found to be effective were 300 µg/ml G418 (Invitrogen) and 0.5 to 1 µg/ml (as indicated in main text) Puromycin (Invitrogen). The transfected cells were grown in G418 containing media for 14-20 days and single colonies were picked and expanded, or in Puromycin containing media for 4 days and harvested as a bulk population and expanded.

2.6. DNA purification

Cells were harvested by brief trypsinization and centrifuged at 300×g for 5 minutes at room temperature. The cells were resuspended in 200 µl PBS, lysed by addition of an equal volume of 2× lysis buffer (100 mM Tris (Sigma) pH8.0, 40 mM EDTA, 2% sodium dodecyl-sulphate (SDS, Biorad), 200 µg/ml Proteinase K) and incubated overnight at 55^{o}C to allow for complete protein digestion. Subsequently, the samples were purified by one round of phenol (P) (Sigma), phenol / chloroform (BDH) / isoamyl-alcohol (Sigma) (25/24/1 v/v) (PCI), chloroform / isoamyl-alcohol (24/1 v/v) (CI) and the purified aqueous phase was incubated with 100 µg/ml RNase A (Roche) for 1 hour at 37^{o}C, before a second round of P-PCI-CI extraction. DNA was dehydrated by addition of 40 µl 5 M NaCl (0.5M final) and 400 µl isopropanol for 5 minutes to overnight, depending on the size of the DNA fragments being purified, at 4^{o}C. The samples were centrifuged at 16,000×g for 5 to 20 minutes at room temperature to pellet the DNA precipitate. The DNA pellets were washed once with 70% ethanol and left to air-dry. After all the ethanol had evaporated, the pellets were dissolved in an appropriate amount of H_2O or 0.1× TE buffer (1 mM Tris pH8.0, 0.1 mM EDTA pH8.0) depending on the downstream analyses. Quantification of DNA samples was performed using a Nanodrop ND1000 instrument (Invitrogen).

2.7. RNA purification

RNA was extracted from cells using either TRIzol Reagent (Invitrogen) or RNEasy spin columns (Qiagen) according to manufacturer's instructions. The purified RNA was quantified on a Nanodrop ND1000 instrument (Invitrogen) and stored at -20^{o}C short term or -80^{o}C long term.

2.8. Reverse transcription and quantitative PCR

Reverse transcription (RT) of 0.5 – 2 µg total RNA was performed using M-MLV reverse transcriptase (Invitrogen), according to the manufacturer's instructions. The reaction was complemented with RNase inhibitors (RNaseOUT, Invitrogen) as suggested by the manufacturer. Real-time quantitative PCR (Q-PCR) was performed using 2× SYBR Green PCR Master Mix (ABI), according to manufacturer's instructions. Samples were amplified either on an ABI 7900HT or an ABI 7500 Real-time PCR instrument. Initial analysis (threshold / baseline setting) was performed using the ABI SDS software. Standard curve quantitation and further analyses of the raw data were performed in Microsoft Excel. Primers used for RT-QPCR expression analyses are shown in *Table 2.2*.

Table 2.2. Primer used for RT-QPCR analyses. All primers were used at 60°C. The *Magoh2* and *Mll2* FW_primary primers were used instead of the FW primer to detect primary transcripts post α-amanitin/DRB treatment.

Primer Name		Sequence
Gapdh	FW	ACCTGCCAAGTATGATGACATCA
	REV	GGTCCTCAGTGTAGCCCAAGAT
Magoh2	FW	CGGGCATAAGGGCAAGTTT
	REV	AATTACTGTTGTTGGCATATGTAAGCTT
	FW_primary	TGGGACTATGAAATTGTCTTTACC
Mll2	FW	GCGTCCACCACATCAAAGTC
	REV	TCGCTCTGTGCTGACTGACA
	FW_primary	ACATCCACTCACCTCAAACCTT
18S rRNA	FW	CAGTAAGTGCGGGCCATAAG
	REV	GGCCTCACTAAACCATCCAA
Magoh	FW	CAGGACCTGAAGTGTTTAGTC
	REV	CTTGTGTCCACAATATCCAAT
Oct4	FW	AGGTGGAACCAACTCCCGAG
	REV	GCTTCAGCAGCTTGGCAAC
Scl/Tal1	FW	CCAACAACAACCGGGTGAAG
	REV	GCCGCACTACTTTGGTGTGAG
Gata2	FW	AGAACCGGAAGCTCATC
	REV	TCGTCTGACAATTTGCACAACAG
Fli1	FW	TCGTGAGGACTGGTCTGTATGG
	REV	GCTGTTGTCGCACCTCAGTTAC
Runx1	FW	GCAGGCAACGATGAAAACTACTC
	REV	CAAACCTGAGGTCGTTGAATCTC
Pu.1	FW	CCATAGCGATCACTACTGGGATTT
	REV	TGTGAAGTGGTTCTCAGGGAAGT
Csf1R	FW	CTTTGGTCTGGGCAAAGAAGAT
	REV	CAGGGCCTCCTTCTCATCAG
Gfi1	FW	GTGAGCCTGGAGCAACACAA
	REV	CTCTTGAAGCTCTTGCCACAGA
Tie2	FW	TGCAACTGAAGAGAGCAAATG
	REV	TCAAGCACAGGATAAATTGTG
Evi1	FW	TGCCCTGGAGATGAGCTGTAA
	REV	GATCTAGAGCAGAAAGGCCAGATT
Cdh5	FW	AATTCTTCCGAATAACCAAGC
	REV	GCACAATGGACTCTTTCCCTA
Icam2	FW	GTGTACCAGCCTCCAGCTC
	REV	CAAAGGTCTGATTCTTCAGGG

2.9. *In vivo* and *in vitro* dimethyl-sulphate treatment

Dimethyl-sulphate (DMS) induces the formation of *N*-7-methyl-guanine (70%) and *N*-3-methyl-adenine (30%) (Brookes and Lawley, 1963). Subsequent piperidine (Fluka) treatment introduces single strand breaks at the sites of the *N*-7-methyl-guanine modification (Maxam and Gilbert, 1977). Applying the same treatment to living cells (*in vivo* DMS treatment) and naked DNA (G-Reaction or *in vitro* DMS treatment) provides information about transcription factor occupancy as detected by either hyper-reactivity or protection of certain guanine residues, visualized on a polyacrylamide gel after ligation mediated PCR (LM-PCR) (Tagoh et al., 2006). Murine ES cells were harvested and washed twice with PBS. 1-2×10^6 cells were resuspended in 100 µl of 0.2% v/v DMS (Aldrich) in PBS and incubated for 5 minutes at room temperature. After DMS treatment the cells were washed twice with 20 ml ice cold PBS. Finally the cells were resuspended in 200 µl of ice cold PBS, lysed by addition of an equal volume of 2× lysis buffer (100 mM Tris pH8.0, 40 mM EDTA, 2% SDS, 200 µg/ml Proteinase K) and incubated overnight at 55^{o}C. Subsequently, the samples were purified as described in paragraph 2.6. The DMS treated genomic DNA was left to dissolve overnight in 99 µl 0.1×TE (1 mM Tris-HCl pH8.0, 0.1 mM EDTA pH8.0). For the G reaction, 100 µg of E14 intact genomic DNA were left overnight to dissolve in 100 µl of H_2O. 100 µl of 2× DMS buffer (100 mM sodium cacodylate pH 8.0, 2 mM EDTA pH8.0) were added to the DNA and mixed by inverting the

tube. 10 µl of 10% DMS were added to the mixture and the samples were incubated for 3 minutes at room temperature. The methylation reaction was stopped by addition of 30 µl DMS stop buffer (1.5 mM sodium acetate pH7.0, 1 M β-mercaptoethanol) and 750µl ice-cold ethanol and allowed to precipitate at room temperature. DNA pellets were washed with 70% ethanol and left overnight to dissolve in 99 µl 0.1× TE. Methylated genomic DNA (*in vivo* or *in vitro*) was then cleaved by piperidine (Fluka). 1 µl of 10 M piperidine was mixed thoroughly with 99 µl of DMS treated sample and was incubated for 10 minutes at 90°C. Samples were cooled in ice cold water and diluted to 400 µl with H_2O. Subsequently, piperidine was removed by extracting twice with 800 µl isobutanol (BDH). Residual isobutanol was removed by a chloroform extraction. The aqueous phase was transferred to a clean reaction tube and the DNA was precipitated by addition of 40 µl 5 M NaCl and 400 µl isopropanol. DNA pellets were resuspended in 0.1× TE buffer, quantified using a Nanodrop ND-1000 spectrophotometer and diluted to 1 µg/µl with 0.1× TE buffer. A small amount of the purified samples was separated on a 0.8% agarose, 0.5× TBE gel to confirm that sufficient piperidine cleavage had occurred.

2.10. *In vivo* and *in vitro* DNaseI / MNase treatment

The method for DNaseI treatment of permeabilised cells is adapted from the method described by Pfeifer and Riggs (Pfeifer and Riggs, 1991). For the *in vivo* DNaseI treatment, ES cells were harvested by trypsinization and washed once with PBS. The cells were counted and resuspended in ice-cold Ψ buffer (11 mM KPO_4 pH7.4, 108 mM KCl, 22 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM DTT) containing 1 mM ATP (freshly added) at a concentration of 3×10^7 cells/ml and were kept on ice. A range of DNaseI concentrations was prepared by mixing 80 µl Ψ buffer containing 1 mM ATP with 4 µl 10% non-ident P40 (NP40) and 20 µl H_2O containing 2 U, 4 U or 8 U DNaseI (Worthington). The 104 µl of the above solution were mixed with 100 µl of the cell suspension and were incubated for 6 minutes at room temperature. The reaction was stopped by addition of 200 µl 2× lysis buffer. The *in vitro* DNaseI controls were generated by initially dissolving 30 µg E14 intact genomic DNA in 100 µl Ψ buffer. A range of DNaseI concentrations was prepared by mixing 80 µl Ψ buffer, 4 µl 10% NP40 and 20 µl H_2O containing 2 U, 4 U or 8 U DNaseI. The 104 µl of the above solution were mixed with the 100 µl of DNA and incubated for 3 minutes on ice. The reaction was stopped by addition of 200 µl 2× lysis buffer. The samples, both *in vivo* and *in vitro*, were incubated overnight at 55^{o}C, purified as described in paragraph 2.6 and then dissolved in 0.1× TE at a concentration of 1µg/µl. Digested DNA was separated on a 0.8% agarose, 0.5× TBE gel to verify DNaseI digestion

and identify the samples most suitable for downstream analyses. For the *in vivo* Micrococcal nuclease (MNase) treatment, ES cells were harvested by trypsinization and washed once in PBS. The cells were resuspended at a density of 3×10^7 cells/ml of Ψ buffer. A range of MNase (Worthington) concentrations was prepared by mixing 80 µl Ψ buffer, 4 µl 10% NP40 and 20 µl H_2O containing 4 U, 8 U or 16 U MNase. The 104 µl of the MNase containing solution were mixed with 100 µl of the cell suspension and incubated for 6 minutes at room temperature. The reaction was stopped by addition of 200 µl 2× lysis buffer. The samples were then incubated overnight at 55^{o}C and purified as described in paragraph 2.6. Finally the DNA pellets were dissolved in 0.1× TE at a concentration of 1 µg/µl. MNase digests were separated on a 0.8% agarose, 0.5× TBE gel to identify the samples most suitable for downstream analyses. Importantly, for both DNaseI and MNase a subset of cells was treated in the exact same way, except there was no nuclease present. This was used to verify that the observed digestion was indeed due to the addition of the nuclease and not due to DNA degradation or apoptosis. Also the 0 U DNaseI treated sample was digested with appropriate restriction enzymes to generate a size marker in the DNaseI hypersensitive site mapping experiments.

2.11. DNaseI hypersensitive site mapping

DNaseI treatment was performed as described in paragraph 2.10. The method used for transfer, probe labelling and hybridization was previously described by Cockerill (Cockerill, 2000). A series of DNaseI treated samples (10 µg each) were digested to completion in 16 µl of 1× the appropriate NEB restriction enzyme buffer ± BSA and 10 U of the restriction enzyme (RE) (NEB) indicated in each figure. Digestions were performed at an appropriate for the enzyme temperature, for 3-4 hours. The RE size marker was created by a complete digestion of 5 µg 0 U DNaseI treated sample with the same enzyme used for the DNaseI treated samples and subsequent partial digestion with REs (indicated on figures) that would generate conveniently sized fragments. The restriction reactions were stopped by addition of 4 µl of 20% ficoll 400 (Sigma), 1% SDS, 0.1% Orange G (Aldrich). The DNaseI digests (10 µg) and the size markers (5 µg) were loaded on a 0.8% agarose, 1× TAE (Severn Biotech) gel containing 0.5 µg/ml ethidium bromide and were separated for 16 hours at 45 V in 1× TAE buffer containing 0.5 µg/ml ethidium bromide. Under these conditions the Orange G dye migrated approximately 20 cm. The digests were then visualised in a Biorad Geldoc station, keeping UV exposure to a minimum. The gel was washed twice for 15 minutes in 0.5 M NaOH, 1.5 M NaCl to denature the DNA and subsequently washed twice for 20 minutes in 1 M Tris pH7.0, 1.5 M NaCl to neutralise the NaOH. The DNA was then transferred to a Biorad ZETA-Probe membrane overnight in

10× SSC (0.15 M sodium citrate (Sigma) and 1.5 M NaCl). The transfer assembly was taken apart and the membrane was washed in 2× SSC and rinsed with H_2O. The DNA was then UV-fixed on the membrane with 0.07 J/cm^2 in a FlowGen UV cross-linker. The membrane was blocked by pre-hybridization for 60-90 minutes with rotation at 65^{o}C in 20 ml RapidHyb buffer (GE healthcare) containing 0.25 mg/ml sonicated and denatured herring testes DNA (Roche). 40 ng of the probe were radiolabelled using the Amersham Megaprime kit (GE Healthcare) and [α^{32}-P] dCTP (Perkin Elmer), according to the manufacturer's instructions (for some probes [α^{32}-P] dATP was used in combination with [α^{32}-P] dCTP to obtain higher labelling efficiency). The labelled probes were separated from unincorporated nucleotides by passing through a Sephadex G50 spin column (GE healthcare), mixed with 5 mg herring testes DNA and denatured at 99°C for 5 minutes. The probe was cooled on ice and mixed with the pre-hybridization buffer. The membrane was left to hybridize for 2 hours with rotation at 65°C. The membrane was then washed twice with 2× SSC, 25 mM sodium phosphate (Sigma), 0.1% SDS at 65°C for 15 minutes and the background radioactivity level (at the corner of the membrane) was estimated with a Geiger counter. The membrane was then washed twice for 20 minutes with solutions of increasing stringency (0.5× SSC, 0.25× SSC, 0.1× SSC), until the background radioactivity was sufficiently low. The membrane was sealed in a plastic film and exposed for at least 16 hours to a Biorad K-Screen. The screen was scanned on a Biorad Pharos FX molecular imager. The size of the DNaseI produced bands, was

estimated by measuring the distance they migrated and relating it to the migration distance of the known size RE marker. The probes used in DNaseI hypersensitive site mapping experiments were generated by nested PCR on E14 genomic DNA and gel purification using the Qiagen QIAquick Gel Extraction Kit according to the manufacturer's instructions. The probe used for *Mll2* genotyping was generated by a single PCR on an *Mll2* BAC template. The primers and PCR conditions used to generate the probes and the genomic co-ordinates of the probes are shown in *Table 2.3*.

Table 2.3. Primers used to generate probes for Southern blots. Also shown on the table, the amplicon length and the genomic region to which the probe is complementary. The *Mll2* probe was generated by a single PCR on an *Mll2* BAC as template.

Primer name		Sequence	Tα°C	Amplicon	Genomic Target
Probe 1	FW	AAGTTAGCCTTTGTTCCCTATGAAATA	65	1047bp	
	REV	ATGAATATAGGTGTCCTTGGGGTAG			
	FW nest.	AAATTGCTGAAACAAAAATGAAACCCT	66	809bp	chr6: 131234803 - 131235611
	REV nest.	CATGTCAGAGCTTGTCTGTTAGCCA			
Probe 2	FW	AGACTACCATGGTGCAAAGGA	60	1044bp	
	REV	CAACCTGAGACCCAAAGGAA			
	FW nes.	TGGTGCAAAGGAGTTAATTTCA	60	633bp	chr6: 131247325 - 131247957
	REV nest.	TTGCCATACGCTGAACACAT			
Probe 3	FW	CACTTTCTCGCCTCTTCACA	60	617bp	
	REV	ACCTCTGAGCCATCTCTCCA			
	FW nest.	CGCCTCTTCACAGGAGATAAA	58	520bp	chr6: 131239967 - 131240486
	REV nest.	TTCTTGTTCCTCTTGCATTTGA			
Probe 4	FW	AACCAATTTGAAAGCACTCAG	55	703bp	
	REV	TGCTGTAAAGTAATTCAGAAAAGC			
	FW nest.	ACCTGCAATTCCTCTCTCCA	58	624bp	chr6:131242072 - 131242695
	REV nest.	TGCCACTCCTGAAAAGTTGA			
Mll2	FW	GTAGTCCAGTCGCCTTGCTC	60	824bp	chr7:31374024 - 31374847
	REV	CGAGACAGGAGCTACGAACC			

2.12. *In vivo* footprinting

2.12.1. Ligation mediated PCR on DNaseI/DMS treated DNA

Ligation mediated PCR (LM-PCR) was performed as described previously (Tagoh et al., 2006) with a few modifications. All primers were designed with theoretically optimal properties (as described in Grange et al., 1997) using the Oligo 5.0 software. The G-reaction DNA was used as a template to optimise the primers' annealing temperatures in LM-PCR. The method is a multi-step procedure, were a biotinylated primer is used for a primer extension reaction, performed in 5 µl of 1× Thermo Pol buffer (10 mM KCl, 20 mM Tris-HCl pH8.8, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100) (NEB), 0.25 µM dNTPs (Promega), 0.20 µM biotinylated 1st primer, 0.3 M sulpholane (Aldrich), 1 U Vent Exo$^-$ polymerase (NEB) and 1 µg template DNA (DMS-piperidine/DNaseI digestion product). The above mixture was initially denatured for 15 minutes at 95^{o}C, the 1st primer was allowed to anneal for 20 minutes and finally, extension was performed at 76^{o}C for 20 minutes. The extension products were ligated to a custom linker (LP25/21 linker) by mixing the 5 µl extension product with 7.25 µl of 23.3% PEG 6000, 48.7 mM Tris pH7.5, 13.5 mM MgCl2, 32.8 mM DTT, 1.65 mM ATP, 0.08 mg/ml BSA, 5.5 µM LP25/21 linker and 5 U T4 DNA ligase (Epicentre) and overnight incubation at 16^{o}C. Streptavidin coated magnetic beads (Dynal M-280, Invitrogen) were used to purify the linker ligated extension products.

Firstly, 10 µl of the stock beads solution was washed twice on a magnetic separator with 2× bind & wash buffer (B&W buffer: 10 mM Tris pH7.5, 1 mM EDTA, 2 M NaCl). The washed beads were resuspended in 12.5 µl 2× B&W buffer and added to the ligation products. The bead/ligation product mixture was incubated at room temperature with rotation for 2-4 hours. The beads were separated from the solution on a magnetic separator and the supernatant was discarded. The beads were washed once with 200 µl 2× B&W buffer, twice with 1× TE buffer and were finally resuspended in 10 µl of 0.1× TE buffer. Subsequently, the bound DNA was released from the beads by incubating at 95^{o}C for 15 minutes. The captured products were amplified using a gene specific primer and a primer specific for the linker (LP25). A PCR master mix was added to the 10 µl of the eluted DNA to a 50 µl final volume of 0.2 mM dNTPs, 0.5 µM 2nd primer, 0.5 µM LP25 primer, 3% dimethyl-sulphoxide (DMSO), 1× Phusion GC buffer (NEB), 1 U Phusion Hot-start polymerase (NEB). Amplification was performed by 30" initial denaturation at 98^{o}C followed by 22 cycles of 10" denaturation at 98^{o}C, 10" annealing, 30" extension at 72^{o}C and finally an extension completion for 5 minutes at 72^{o}C. The 3rd primer was radioactively labelled by a 1 hour incubation of a 10 µl solution of 4 µM 3rd primer, 1× polynucleotide kinase buffer (PNK buffer: 70 mM Tris pH7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT)), containing 5 µl of EasyTide [γ^{32}-P] dATP (Perkin Elmer) and 10 U T4 polynucleotide kinase (NEB) at 37^{o}C and subsequently diluted to 50 µl with H_2O and purified by passing through a G25 resin spin column (GE Healthcare). 10 µl of the

amplification product were mixed with 4 µl of the labelling mix (0.25 mM dNTPs, 3.75% DMSO, 1× Phusion GC buffer, containing 0.5 U Phusion Hot-Start polymerase and 2 µl of the radio-labelled 3rd primer) and radio-labelled by primer extension (30" initial denaturation at 98°C followed by 7 cycles of 10" denaturation at 98°C, 10" annealing, 30" extension at 72°C and finally an extension completion for 5 minutes at 72°C). All PCR reactions were performed in a Biometra T3000 Thermal Cycler. Primer sequences and PCR conditions are shown in *Table 2.4*.

2.12.2. LM-PCR on Micrococcal nuclease treated DNA

The method used for MNase footprinting is slightly different to the one employed for DMS/DNaseI, as MNase catalyses double strand cuts and MNase cleaved DNA lacks the 5'-phosphate group, thus being unsuitable for ligation. Initially, 10 µg of MNase treated DNA were phosphorylated by a 1 hour incubation in a 30 µl reaction volume of 1× PNK buffer, 1 mM ATP and 20 U T4 polynucleotide kinase, at 37°C. Volume was adjusted to 50 µl to yield a DNA concentration of 0.2 µg/µl. Two micrograms of the phosphorylated DNA were ligated to 100 pmol of the LP25/21 linker in 27 µl of 30 mM This pH7.5, 8.5 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 0.05 mM BSA, containing 6 U of T4 Ligase. The ligation reaction was performed overnight at 16°C. After ligation, the final volume was adjusted to 100 µl with 0.1× TE and the ligation products were

precipitated by addition of 10 µl 5M NaCl and 100 µl isopropanol. The samples were centrifuged at 16,000×g for 10 minutes at 20^{o}C and the DNA pellets washed once with 70% ethanol. DNA pellets were dissolved in 0.1× TE to give a final concentration of 1 µg/µl. 1µg of the linker-ligated product was used in a primer extension reaction and the products were captured on streptavidin coated magnetic beads (as described for LM-PCR on DMS/DNaseI treated material in paragraph 2.12.1). Subsequently, the captured products were eluted in 10 µl 0.1× TE by heating to 95^{o}C for 15 minutes. The whole eluate was used in an amplification reaction using gene specific and linker specific primers (as for DNaseI/DMS). Finally, labeling of the amplified products was performed as described above.

2.12.3. DNA denaturing polyacrylamide gel electrophoresis

The radiolabelled LM-PCR products were mixed with 14 µl of 2× sequencing gel loading dye (0.1% w/v bromophenol blue, 0.1% w/v xylene cyanol in formamide). DNA was denatured by heating at 95^{o}C for 10 minutes and cooled in ice cold water. LM-PCR products were separated on a 6% polyacrylamide (19:1 acrylamide:bis-acrylamide) 8M urea gel prepared using the UreaGel Sequencing System (GeneFlow), according to the manufacturer's instructions. The gel was allowed to polymerize for at least one hour at room temperature. The gel was pre-ran

in a Gibco S2 sequencing apparatus at 85 W for approximately 90 minutes to reach a temperature higher than 50°C. The denatured LM-PCR products were loaded on the gel and were ran for 90 minutes at 85 W. The gel was allowed to cool before separating the glass plates and was then fixed for 2 minutes in fixing solution (10% methanol (Sigma), 10% acetic acid (Sigma) in H_2O). The gel was transferred to a sheet of Whattman 3MM paper, covered with cling-film and dried in a prewarmed Biorad gel dryer for 30 to 90 minutes. Images were obtained by overnight exposure of a Biorad K-Screen to the dried gel. The screens were scanned using a Biorad Pharos FX molecular imager.

Table 2.4. Primers used for footprinting experiments. Primers VL1 and 13 target different regions of the Magoh2 CpG Island promoter. Primer VL2 is in reverse orientation to VL1 and 13. The *Oct4* primers were used to verify that the samples used were equally digested.

Primer name		**Sequence**	**Tα°C**
VL1	a	[bio]-TTCCCTACCGGAGTAAAC	50
	b	TGCGCCCGTGACGTCACTACC	58
	c	TCACTACCTCGCGCCGGCGGC	66
VL2	a	[bio]-CACGCTGAAACCCGCTC	50
	b	CTCACCGTCCGGCCGAAACT	58
	c	GCCGAAACTCGAACTCCAAAAACTC	64
VL13	a	[bio]-AGCGCTCGAGATCGTGGT	58
	b	GTGGCTCCAATGCGAACCTTCAG	66
	c	GGCTCCAATGCGAACCTTCAGTTCTCT	68
Oct4	a	[bio]-CAACTGGTTTGTGAGGTGTCC	52
	b	GTGACCCAAGGCAGGGGTGAGA	60
	c	CCCAAGGCAGGGGTGAGAGGACCT	62

2.13. Measurement of DNA methylation

2.13.1. Measurement by bisulphite conversion and pyrosequencing

Genomic DNA was treated with sodium bisulphite using the EZ-DNA methylation kit from Zymo Research according to the manufacturer's instructions. Bisulphite converted DNA was used as a PCR template to amplify the *Magoh2* promoter using primers specific for the converted DNA sequence (*Table 2.5*). The PCR products were sent to the pyrosequencing facility in the University of Leeds where they were purified using streptavidin coated magnetic beads and prepared for sequencing using a third sequencing primer (*Table 2.5*).

2.13.2. Measurement by methylation sensitive restriction enzyme digestion and QPCR

500 µg of genomic DNA were digested to completion with 10 U HpaII in a final volume of 10 µl. The digests were diluted to an estimated concentration of 2 ng/µl and used as a template for a QPCR reaction with primers flanking a single HpaII site on the *Magoh2* promoter (*Table 2.5*). The data were subsequently normalised against a PCR amplicon on the *Oct4* promoter (*Oct4 primers shown in Table 2.8*), that does not contain a HpaII site.

Table 2.5. Primers used to measure DNA methylation. BiS indicates primers used for bi-sulphite converted DNA. 2 primer sets had to be used to cover the whole Magoh2 CpG island (5' and 3').

Primer name		Sequence	Tα°C
BiS Magoh2 5'	FW	AGTAGAGAAGGTAGAAATTATTATTTATAG	53
	REV	[bio]-CTCTTAAAAATCTCTTACTTCCTCTTC	
	SEQ	GGTAGAAATTATTATTTATAGATAT	
BiS Magoh2 3'	FW	AGGGTAAAAATGTTTATGGGTAGT	56
	REV	[bio]-AAACCCTTTCTCAAAACTAAACAAAT	
	SEQ	GGTATAAGGGTAAGTTTGG	
Magoh2 HpaII	FW	TTCTCTTGGGGGTCTCTTGCTTCC	60
	REV	CGCGTCACCAAGGGCGCGTT	

2.14. Protein purification

2.14.1. Whole cell protein extracts

Cells where harvested by scraping in an appropriate volume of cold PBS and centrifuged at 300×g for 5 minutes at 4°C. The cells were resuspended in residual volume, transferred in 1.5 ml tube and centrifuged again. After all the supernatant was removed, the cells were lysed by addition of 1 µl RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH8.0) per 5×10^5 cells and incubated for 10 minutes on ice. The lysate was centrifuged at 16,000×g for 10 minutes at 4°C and the supernatant was transferred to a clean tube. 5× reducing disruption buffer (250 mM Tris pH6.8, 0.5 M DTT, 50% glycerol, 25% β-mercaptoethanol, 10% SDS, 0.1% bromophenol blue) was added to the protein extract and mixed thoroughly. Samples were stored at -20°C.

Alternatively, total protein extracts were obtained by acetone precipitation of lysates from cells used for RNA purification by the Qiagen RNEasy spin columns, according to the manufacturer's instructions. The precipitated proteins were dissolved in urea buffer (40 mM Tris pH6.8, 8 M urea) and then centrifuged at 16,000×g for 10 minutes at 4^{o}C. The supernatant was transferred to a clean tube, mixed thoroughly with 5× reducing disruption buffer and stored at -20^{o}C.

2.14.2. Nuclear protein extracts

Cells were harvested by brief trypsinization and spun down for 5 minutes at 300×g at 4^{o}C. The cells were washed in 1 ml PBS and resuspended in sucrose buffer (0.32 M sucrose, 50 mM KCl, 20 mM NaCl, 3 mM $CaCl_2$, 2 mM magnesium acetate, 10 mM Tris pH8.0, 1 mM DTT, 0.5 mM phenylmethylsulphonyl fluoride (PMSF), 1.5 mM spermine, 0.5 mM spermidine, 1:1000 protease inhibitor cocktail (PIC: 104 mM AEBSF, 80 µM aprotinin, 4 mM bestatin, 1.4 mM E-64, 2 mM leupeptin, 1.5 mM pepstatin A, Sigma) at a density of 2×10^8 cells/ml. An equal volume of sucrose buffer containing 0.2% NP40 was added to disrupt the plasma membranes. Intact nuclei were spun down for 5 minutes at 500×g/4^{o}C and washed with 1 ml sucrose buffer (without NP40). The nuclei were resuspended in half the volume of sucrose buffer used initially (minimum of 20 µl) of low salt buffer (10 mM 4-(2-hydroxyethyl)-1-

piperazine-ethane-sulphonic acid (HEPES) pH8.0, 20% glycerol, 2 mM magnesium acetate, 20 mM KCl, 10 mM NaF, 1 mM sodium pyrophosphate, 2 mM ethylene glycol tetra-acetic acid (EGTA), 0.5 mM DTT, 0.5 mM PMSF, 1:1000 PIC). An equal volume of high salt buffer (10 mM HEPES pH8.0, 20% glycerol, 2 mM magnesium acetate, 0.7 M KCl, 10 mM NaF, 1 mM sodium pyrophosphate, 1% NP40, 2 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF, 1:1000 PIC) was added drop-wise, while slowly mixing the lysate. The lysates were incubated on ice for 20 minutes and then spun down at 16,000×g for 5 minutes at 4^{o}C. The supernatant was passed through an ion exchange column (Thermo Fisher) and eluted in buffer D (20% glycerol, 10 mM HEPES pH8.0, 50 mM NaCl, 45 mM KCl, 2 mM magnesium acetate, 10 mM NaF, 1 mM sodium pyrophosphate, 2 mM EGTA, 0.1 mM $ZnCl_2$, 1 mM DTT). The purified nuclear extracts were stored at -80^{o}C.

2.15. Electrophoretic mobility shift assays

2.15.1. Probe generation

To generate double stranded (ds) DNA probes for electrophoretic mobility shift assays (EMSAs), custom complementary single stranded (ss) DNA oligonucleotides were ordered from Sigma. The oligonucleotides (500 pmol each) were annealed to form dsDNA probes in 100 µl 1× oligo

annealing buffer (20 mM Tris pH8.0, 0.2 mM $MgCl_2$, 1 mM EDTA pH8.0, 50 mM NaCl). The mixture was heated to 100^{o}C on a heat block for 15 minutes and was then left to slowly cool down to room temperature on the block. The annealed probes were mixed with glycerol (10% final) and were separated on a 0.5× TBE, 18.5% glycerol, 7% polyacrylamide gel for approximately 2 hours at 150 V. The gel was then stained in 0.5× TBE containing 0.5 µg/ml ethidium bromide for approximately 20 minutes and the dsDNA probes were excised from the gel on a UV transilluminator, minimizing UV exposure as much as possible. The gel slices were homogenized using a clean pair of forceps and transferred to 1.5 ml tubes. The probes were eluted by an overnight incubation in 750 µl oligo elution buffer (20 mM Tris pH8.0, 400 mM NaCl, 2 mM EDTA pH8.0, 0.05% SDS) at 37^{o}C. The eluted probes were passed through a 0.45 µm syringe filter to remove polyacrylamide pieces and then purified using standard phenol extraction as described in paragraph 2.6.

2.15.2. Protein binding and band shift assay

4 µg of nuclear extract in buffer D (described in 2.13.2) were mixed with the non-labelled competitor and 1 µg antibody (where indicated) for 15 minutes at room temperature. The labelled probe was added to the reaction and incubated for a further 25 minutes at room temperature. The quantity of labelled probe used was calculated to give 20,000 cpm (1 to 15

pmoles). The amount of non labelled competitor used was 100× the amount of the respective probe. The samples were separated at 150 V for 1 hr 15 minutes on a 5% acrylamide gel (5% acrylamide (37.5:1 Seven Biotech), 0.5× TBE (Seven Biotech), 1/1000 tetra-methyl-ethylene-diamine (TEMED, Fluka) and 0.1% ammonium persulphate (APS, Sigma) that was allowed to set for 1 hour before pre-run at 150 V in 0.5× TBE for 60 minutes. The gel was fixed with 10% methanol, 10% acetic acid for 5 minutes and dried in a gel dryer for 90 to 120 minutes. Biorad K-Screens were exposed to the gel overnight and scanned on a Biorad PharosFX molecular imager.

2.16. SDS polyacrylamide gel electrophoresis and Western blotting

Whole cell extracts or nuclear extracts from equal amounts of cells were separated on a Biorad 4–20% Mini-PROTEAN TGX Precast Gel for 1 hour at 200 V in 1× SDS running buffer (24.8 mM Trizma Base (Sigma), 0.19 M glycine, 0.1% w/v SDS). Proteins were then transferred onto a nitrocellulose membrane by wet transfer at 100 V for 70 minutes at 4^{o}C in 24.8 mM Trizma Base, 0.19 M glycine, 10% v/v methanol. The membrane was then washed briefly with Tris buffered saline – Tween20 (TBS-T: 50 mM Tris pH7.6, 150 mM NaCl, 0.05% Tween20) and blocked with 5% w/v bovine serum albumin (BSA, Sigma) in TBS-T for 2 hours at room temperature or 4^{o}C overnight with mild shaking. The primary antibody was added to the blocking

buffer and incubated for 2-4 hours at room temperature with mild shaking. The membrane was washed twice for 2 minutes with TBS-T and then incubated with the appropriate horseradish peroxidise (HRP) conjugated secondary antibody, diluted in 5% BSA in TBS-T for 30 to 60 minutes. The membrane was washed 4 times for 5 minutes with TBS-T and excess buffer was removed. West-pico chemiluminescence (Thermo Fisher) was used to visualize the results according to the manufacturer's instructions. ECL Hyperfilm (GE Healthcare) was exposed to the membranes and developed in Konica Minolta SRX-101A X-Ray film processor. The antibodies and dilutions used are shown in *Table 2.5*.

Table 2.6. Antibodies and dilutions used for western blots.

Target Protein	Host	Conjugate	Vendor	Cat. Num.	Dilution
RNA PolII CTD	Mouse (mAb)	none	abcam	ab5408	1:1000
RNA PolII phospho-S2	Rabbit (pAb)	none	abcam	ab5095	1:1000
NUP188	Rabbit (pAb)	none	abcam	ab86601	1:500
mouse IgG	Goat (mAb)	HRP	Jackson	115-035-174	1:10000
rabbit IgG	Mouse (mAb)	HRP	Jackson	211-032-171	1:10000

2.17. Chromatin immunoprecipitation

2.17.1. Formaldehyde crosslinking

ES cells were grown under standard conditions as described in 2.2.1. Formaldehyde was added to the culture medium (16% Thermo Scientific) to yield 1% final concentration. The cells were incubated with the formaldehyde for 12 minutes at room temperature with mild shaking. The medium was discarded and the cells washed twice with ice-cold 0.4 M glycine in PBS, to quench any residual formaldehyde. The cells were then harvested by scraping in 10 ml 0.5× Trypsin/EDTA and transferred to a 25 ml universal tube (Corning). The trypsin was inactivated by addition of 1ml FCS and the cells were centrifuged at 300×g for 5 minutes at 4^{o}C.

2.17.2. Chromatin purification and sonication

The cells were washed once in ice-cold PBS, resuspended in ice-cold buffer A: 10 mM HEPES pH8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1:1000 PIC and 0.1 mM PMSF and incubated for 10 minutes at 4^{o}C with rotation. At this point 10 µl of the cell suspension were transferred to a haemocytometer to measure the number of cells harvested. The cells were centrifuged at 300×g for 5 minutes at 4^{o}C, resuspended in ice-cold buffer B: 10 mM HEPES pH8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01%

Triton X-100, 1:1000 PIC and 0.1 mM PMSF and incubated for 10 minutes at 4^{o}C with rotation. The isolated nuclei were then centrifuged at 400×g for 5 minutes at 4^{o}C and resuspended in SDS-IP buffer: 25 mM Tris pH8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.25% SDS, 1:1000 PIC and 0.1 mM PMSF, at a density of 5×10^{6} cells/300 µl SDS-IP buffer. The lysed nuclei were transferred to appropriate reaction tubes, depending on the volume and sheared using a Diagenode Bioruptor at the high setting for 10 minutes of 30" sonication pulse, 30" pause at 4^{o}C. The sonicated samples were centrifuged at 16,000×g for 5 minutes at 4^{o}C if the sonication was performed in 1.5 ml tubes or at 4,000×g for 15 minutes if they were sonicated in 15 ml tubes, to clear the sonicated chromatin from insoluble membrane fragments and other cell components. The cleared supernatant was transferred to a clean reaction tube and mixed with 2 volumes of glycerol-IP buffer: 25 mM Tris pH8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 7.5% glycerol, 1:1000 PIC and 0.1 mM PMSF. Finally, the sonicated chromatin was distributed in 300 µl aliquots (chromatin from 1.67×10^{6} cells per aliquot), snap frozen in liquid nitrogen and stored at -80^{o}C.

2.17.3. Immunoprecipitation

10 µl per immunoprecipitation of protein G coated Dynabeads (Invitrogen) were washed twice in citrate/phosphate buffer pH8.0: 24.5 mM citric acid, 76.7 mM NaH_2PO_4 and resuspended in 10 µl citrate/phosphate buffer. Bovine serum albumin (BSA, Sigma) was added to the beads at a final concentration of 0.5% w/v to block non-specific protein binding. An appropriate antibody was added to the beads which were incubated for 2 hours at 4^{o}C with rotation. The antibodies used are presented in *Table 2.6*. One aliquot of chromatin per immunoprecipitation was thawed slowly on ice and added to the 10 µl of the antibody bound Dynabeads. The chromatin was incubated with the beads for 2-3 hours at 4^{o}C with rotation. The reaction tubes were placed on a magnetic separator to capture the beads and discard the supernatant. The beads were resuspended in ice-cold wash buffer 1: 20 mM Tris pH8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 150 mM NaCl and incubated for 5 minutes at 4^{o}C with rotation. The beads were then separated again and washed twice for 5 minutes with ice-cold wash buffer 2: 20 mM Tris pH8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 500 mM NaCl. The beads were captured on the magnetic separator and washed for 5 minutes with ice-cold LiCl buffer: 10 mM Tris pH8.0, 1 mM EDTA, 0.25 M LiCl, 0.5% NP40, 0.5% sodium deoxycholate (Sigma). Subsequently, the beads were washed twice for 2 minutes with ice-cold TE/NaCl buffer: 10 mM Tris pH8.0, 1 mM EDTA, 50 mM NaCl. The beads were finally resuspended in 100 µl freshly

prepared ChIP elution buffer: 0.1 M $NaHCO_3$, 1% SDS, 200 mM NaCl containing 200 µg/ml Proteinase K and incubated overnight at 65°C to reverse the crosslinks and to elute and deproteinate the bound DNA. Input controls were prepared by treating 30 µl of chromatin with 10 µg RNase A for 1 hour at 37°C. Subsequently, 50 µg Proteinase K were added and SDS to yield 0.2% final concentration and incubated at 65°C overnight. Both input and immunoprecipitated DNA were then purified using the Agencourt Ampure PCR purification kit according to the manufacturer's instructions and eluted in 200 µl of 0.1× TE buffer.

Table 2.7. Antibodies used for chromatin immunoprecipitation

Target	Vendor	Cat. Num.	amount/IP
RNA polymerase II CTD	abcam	ab5408	1µg
H3K4me_3	Millipore	07-473	2µl
H3K9ac	abcam	ab4441	1µg

2.17.4. Quantitative PCR and data analysis

The immunoprecipitated DNA was used as template for QPCR on an ABI 7900HT or ABI 7500 real time PCR instrument. 5 µl of the immunoprecipitated DNA were used as template in a 20 µl reaction containing 0.25 µM of each primer in 1× ABI SYBR Green Master Mix. The initial data analysis (threshold/baseline) was performed by the ABI SDS software and further analyses were performed in Microsoft Excel. Standard curve quantitation was used for all the ChIP experiments. Moreover, all the histone modification pull-downs were corrected for

nucleosome density by calculating the enrichment of the given modification over an H3 immunoprecipitation on the same sample. RNA polymerase II immunoprecipitations were corrected for input. The final values presented are obtained by normalizing against a genomic region that should be positive for the given modification/protein and presented either as enrichment over that control region or as fold change over the first/untreated time-point in a time-course experiment. The primers used in ChIP-QPCR assays are shown in *Table 2.8.*

Table 2.8. Primers used in ChIP-QPCR analyses

Primer Name		Sequence
En1	FW	GGAAGTCAAACCCCTCTACTG
	REV	GGATAAGCCACGGCTCAGA
-20bp *Magoh2*	FW	TACTCCGGTAGGAACGAAA
	REV	CTCCAATGCGAACCTTCAGT
+60bp Magoh2	FW	CAAAAACTCGTGCCCAAACT
	REV	CCGAGGGTAAAAATGTCTATG
Oct4 promoter	FW	TGGGCTGAAATACTGGGTTC
	REV	TTGAATGTTCGTGTGCCAAT
Chr2	FW	AGGGATGCCCATGCAGTCT
	REV	CCTGTCATCAGTCCATTCTCCAT

2.18. Bacterial cultures and BAC/plasmid purification

2.18.1. Bacterial strains

The bacterial strains used were all $rpsL^{mut}$ (resistant to streptomycin), $RecA^{-}$ (no endogenous recombinases) *E.coli* DH10β derivatives. The EC1000-pir116 strain expresses the pir116 protein which is necessary to propagate the pR6K plasmids used. The EC1000-pir116 strain was the host used for all pR6K plasmids. The EC1000-pir166 + pSC101-BAD-ETgA-tet, is the same strain only transformed with an L-arabinose inducible construct that expresses the RecET system of recombinases of the Rac prophage, used for linear-linear recombineering and confers resistance to tetracycline. The bacterial artificial chromosome (BAC) host was the GB05-Red strain. These bacteria contain a genomic integration of the Red operon from phage λ under the control of the L-arabinose inducible BAD promoter. The Red operon expresses the Red α,β,γ system for linear into circular recombineering.

2.18.2. *E. coli* solid cultures

Bacteria were streaked on 10cm Petri dishes containing low salt Luria-Bertani (LB) medium (Lennox) pH8.0, complemented with 1.5% w/v agar and the appropriate antibiotic. Antibiotic concentrations used are shown in *Table 2.9.* Bacteria were grown at 37^{o}C, except for strains containing a pSC101 plasmid which can only be propagated at 30^{o}C.

2.18.3. *E. coli* liquid cultures

Liquid cultures were inoculated with a single colony picked from a solid culture. The bacteria were grown in low salt LB pH8.0 complemented with the appropriate antibiotics (see *Table 2.9* for antibiotic concentrations) at 37^{o}C or 30^{o}C (for pSC101 containing strains) with shaking.

2.18.4. Small scale BAC and plasmid preparations

Small scale preparations were performed using a Qiagen Miniprep Kit. Initially, the bacterial culture was centrifuged for 1 minute at 11,000×g and the supernatant was discarded. The cells were resuspended in 200 µl buffer P1 (containing RNase A as per the manufacturer's instructions) and mixed for 10 minutes at 1300 rpm on an Eppendorf thermo-mixer at room temperature. Subsequently, 200 µl of buffer P2 were added to the cell suspension and mixed gently but thoroughly by inverting the reaction tube. Finally, 200 µl of buffer P3 were added and mixed gently but thoroughly by inverting the reaction tube to neutralise the alkaline lysate. The lysate was centrifuged for 20 minutes at 16,000×g to separate proteins, the bacterial genome and other insoluble cellular components. The supernatant was transferred to a clean reaction tube. The supercoiled DNA was precipitated by addition of 600µl isopropanol, incubated for 10 minutes and centrifuged at 16,000×g for 20 minutes. The supernatant was discarded and the DNA pellet washed with 70% ethanol. The ethanol was discarded and the pellet left to air-dry briefly. Finally, the DNA was dissolved in an appropriate volume of H_2O.

2.18.5. Large scale BAC and plasmid preparations

Large scale plasmid preparations were performed using a Qiagen Maxi-prep kit according to the manufacturer's instructions. The kit is unsuitable for bacterial artificial chromosome (BAC) preparations due to the large size of the BAC constructs that would not be released efficiently from the DNA binding matrix in the kit's columns. A different method was used to purify large amounts of BAC DNA. BAC hosts were grown in 250 ml low salt LB with antibiotics for approximately 16 hours. The whole culture was harvested and centrifuged at 4000×g for 10 minutes at 4^{o}C. The cells were resuspended in 10 ml ice-cold solution A (50 mM glucose, 25 mM Tris pH8.0, 10 mM EDTA pH8.0) and lysed using 20 ml solution B (0.2 M NaOH, 1% SDS). Subsequently the lysate was neutralized by addition of 15 ml solution C (3 M potassium acetate, 12% v/v glacial acetic acid) and centrifuged at 8,000×g for 5 minutes to pellet most of the insoluble fraction. The supernatant was carefully removed and filtered through Whatman paper. 30 ml ice-cold isopropanol were added to the cleared lysate, mixed and centrifuged at 4,000×g for 5 minutes at 4^{o}C. The DNA/RNA pellet was washed once with 70% ethanol and resuspended in 3.75 ml TE. Addition of 5 ml ice-cold 5M LiCl and centrifugation at 4,000×g for 5 minutes at 4^{o}C caused the RNA to precipitate. The supernatant (~9 ml) was transferred to a clean tube, mixed with 18 ml absolute ethanol and centrifuged at 10,000×g for 10 minutes at 4^{o}C. The DNA pellet was washed with 70% ethanol and dissolved in 900 µl TE. 20 µg RNase A were added and the mix was

incubated for 1 hour at 37°C to digest any residual RNA. The DNA was mixed thoroughly with 0.45 ml of PEG/NaCl solution (2.5 M NaCl, 20% polyethylene-glycol 6000) and incubated on ice for 15 minutes. The mix was centrifuged at 16,000×g for 10 minutes at 4°C, generating a transparent gelatinous pellet. The supernatant was discarded and the pellet re-dissolved in 500 µl TE. BAC DNA was purified further by standard phenol/chloroform extraction as described in paragraph 2.6

Table 2.9. Antibiotic concentrations used for selection of bacteria

Antibiotic	**µg/ml**	**application**
ampicillin	50	liquid
	100	solid
chloramphenicol	10	liquid
	15	solid
tetracyclin	3	liquid
	5	solid
gentamicin	1	liquid
	2	solid
streptomycin	100	liquid
	200	solid
blasticidin S	30	liquid (solid BAC)
	40	solid

2.19. Generation of an inducible knock-in *Mll2* BAC by recombineering

2.19.1. Experimental design and computer software

This recombineering exercise was aimed at modifying a pre-existing *Mll2*-GFP BAC illustrated in *Figure 2.2*. A gene-trapping stop cassette would be inserted in intron1 to prevent *Mll2* expression from the BAC. However, the cassette would be flanked by loxP sites, allowing for excision of the stop cassette by activation of the Cre-ERT2 recombinase in the host *Mll2*$^{Fc/Fc}$ ES cells and thus re-expression of *Mll2*. The main purpose of this would be to examine whether re-expression of *Mll2* could rescue the *Magoh2* silencing phenotype (*described in Results*). Moreover, the fact that the BACs would be inactive until they were induced would provide the opportunity to establish stable cell lines, activate the BAC transgenes and follow the kinetics of *Magoh2* reactivation. This exercise was designed together with Dr. Jun Fu, during the author's 3 month placement in the Stewart lab (BIOTEC, TU Dresden, Germany). All the *in silico* manipulation and planning of the experiment was performed using the TextCo Gene Construction Kit.

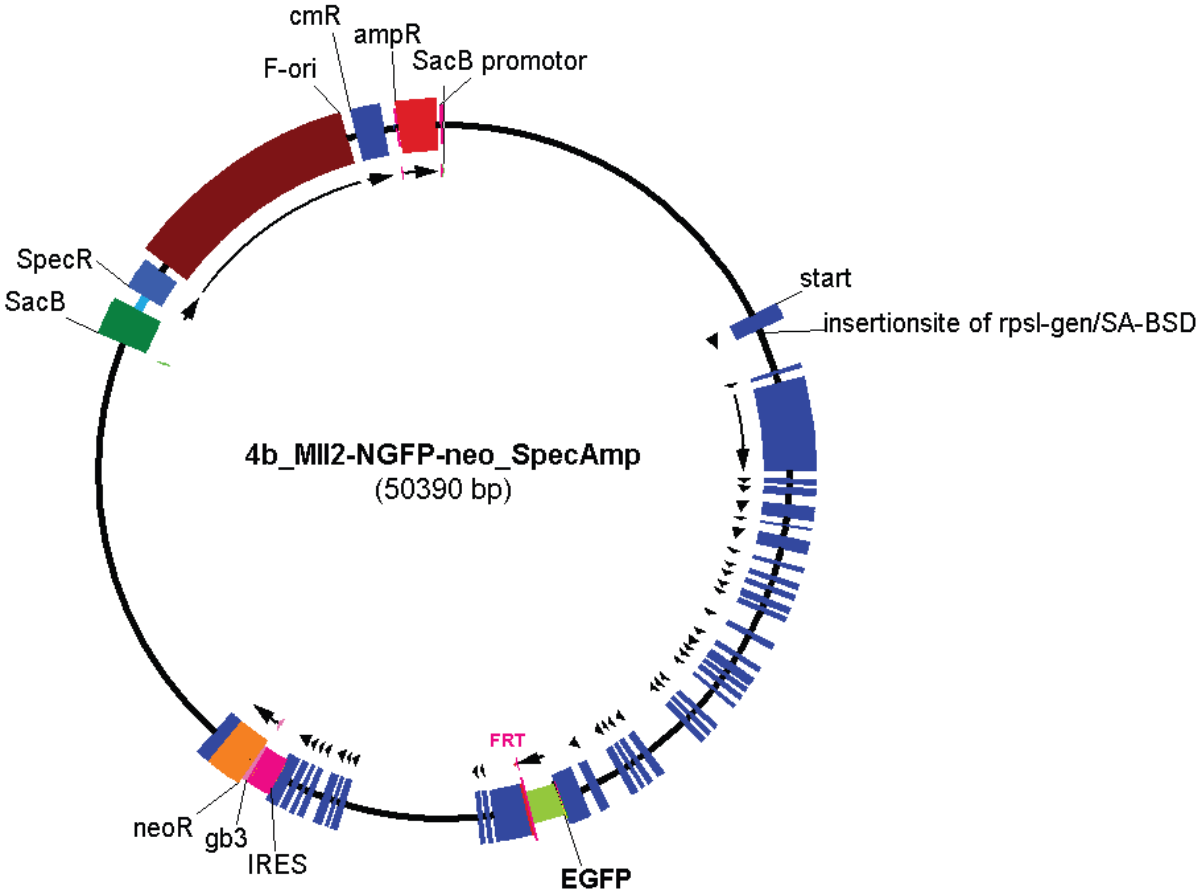

Figure 2.2. Map of the original *Mll2*-GFP BAC (as supplied by Dr. H. Hofemeister in the Stewart lab). F-Ori: BAC origin of replication, cmR: chloramphenicol resistance gene, ampR: ampicilin resistance gene, start: MLL2 translation start site, IRES: internal ribosome entry sequence, gb3: bacterial promoter, neoR: neomycin resistance gene, SacB: bacterial promoter, SpecR: spectinomycin resistance gene, the unlabelled blue boxes are the individual exons of *Mll2*.

2.19.2. Generation of a gene-trapping stop cassette

The gene-trapping cassette was generated by modification of a pre-existing cassette (pR6K-cm-GT0-lacZneo-CoTC) in the Stewart lab, illustrated in *Figure 2.3*. The lacZneo part of the cassette was exchanged for a blasticidin S resistance gene, under the control of the gb3 bacterial promoter (gb3-Bsd). The gb3-Bsd fragment was obtained from the pR6K-2Ty1-2PreS-tdKatushka-biotin-T2A-gb3-Bsd plasmid (*Figure 2.4 top panel*). The pR6K-2Ty1-2PreS-tdKatushka-biotin-T2A-gb3-Bsd was digested to completion with AvaI and the ~1.5 kb double band was gel purified and used as a template for a PCR reaction with primers BSD1 and BSD2 generating the PCR product shown in the lower panel of *Figure 2.4*. The sequences of the primers used for recombineering are shown in *Table 2.9*. The 3' portion of these primers is complementary to the pR6K-2Ty1-2PreS-tdKatushka-biotin-T2A-gb3-Bsd plasmid while the 5' portion is homologous to the pR6K-cm-GT0-lacZneo-CoTC. The homologous regions are labelled in *Figures 2.3* and *2.4* as HA1 and HA2. EC1000-pir116+pSC101-ETgA-tet were cultured overnight at 30^{o}C in tetracycline containing LB (tet-LB). 30 µl of the overnight culture were used to inoculate 1.5 ml tet-LB and were cultured for a further 2 hours on a thermo-mixer at 30^{o}C at 900 rpm. L-arabinose was added to the bacterial cultures (0.2% final concentration) to induce expression of the RecET recombinases and cultured for a further 1 hour at 37^{o}C. The cells were centrifuged at 10,000×g for 1 minute, washed twice with ice-cold H_2O and resuspended in residual volume (~30 µl). 100 ng of the PCR product

and 0.5 µg of SphI/PvuI digested pR6K-cm-GT0-lacZneo-CoTC were added to the bacteria and the mixture was transferred in a pre-cooled 1 mm gap electroporation cuvette. The bacteria were transformed by electroporation at 1350 V. Warm LB culture medium without any antibiotics was added immediately after electroporation and the transformed bacteria were transferred in reaction tubes where they were cultured for 1 hour at 37^{o}C with shaking. Finally, the transformed bacteria were centrifuged at 10,000×g for 1 minute and streaked on LB-agar plates containing 30 µg/ml blasticidin S (Invivogen). Importantly, the transformation was also performed with cells that had not been induced with L-arabinose, to exclude acquired resistance in the absence of recombination. The plates were incubated overnight at 37^{o}C and single colonies were picked and used to inoculate LB medium containing 30 µg/ml blasticidin S. The colonies were grown overnight at 37^{o}C with shaking and the plasmid contained was extracted using the method described in paragraph 2.17.4. The isolated plasmids (10-20 µg) were digested to completion with 10 U PvuII and resolved on an agarose gel. The PCR product would introduce a novel PvuII site in the plasmid therefore enabling detection of successful recombineering. Two clones with the correct digestion pattern were sent for sequencing (core sequencing facility, Max Planck Institute, Dresden, Germany) to verify the integrity of the construct.

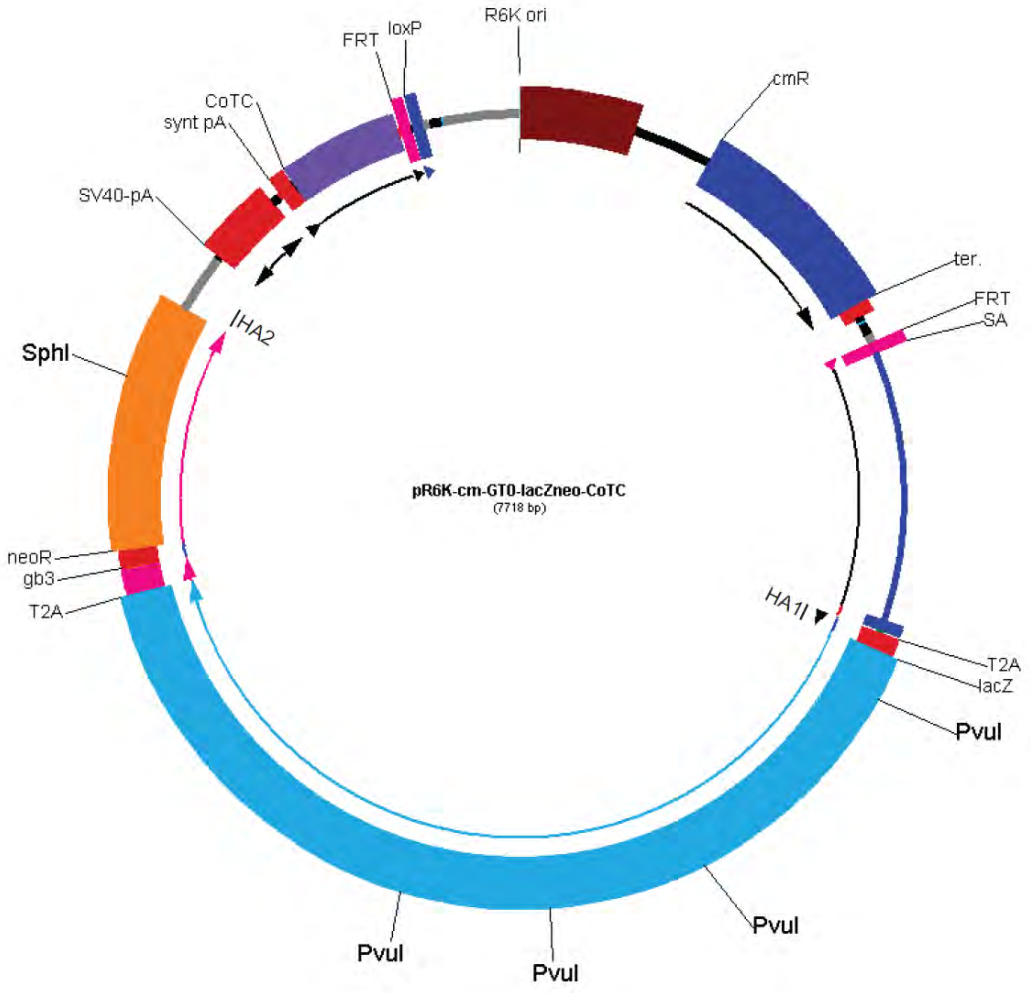

Figure 2.3. The pR6K-cm-GT0-lacZneo-CoTC plasmid. This was used as a base to generate a STOP cassette. The regions labelled HA1 and HA2 were used to swap lacZ-neo cassette for a gb3-Bsd, thus making the construct smaller and easier to work with. R6K Ori: origin of replication, cmR: chloramphenicol resistance gene, FRT: site recognised by the Flp recombinase, SA: splice acceptor, GT0: adapter sequence to maintain the T2A in-frame, T2A: sequence that instructs the ribosome to cleave the nascent peptide and begin synthesizing a new one, SV40-pA: poly-A signal from the SV40 virus, synt-Pa: poly-A signal, CoTC: transcription termination sequence, loxP: site recognised by the Cre recombinase. PvuI and SphI mark the sites for the respective restriction enzymes.

Table 2.10. Sequence of primers used in recombineering.

Primer	Sequence
Mll2-genta1	GTTGTGGACAAGTGGGATTCTCGGGGCTCCTGGAAAGAAGATAACT TCGTATAGCATACATTATACGAAGTTATGCGTGTTTCGAGCATGTT TCTGCGTAGTGTCAGCTCATCC
Mll2-genta2	TAGCTACCCAAACGCTCCACGCAGCCTGCTTCCTTCGTGGTGTCCA TCATCCTGTAGGTGTAGACGACGACGAACAGAG
BSD1	TCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTGCTATG GCCACAACCTGTTAC
BSD2	TCTGGTTATGTGTGGGAGGGCTAAGCGGGACTCTGGGGTTCGAAAT GACCGACCAAGCGACGCC

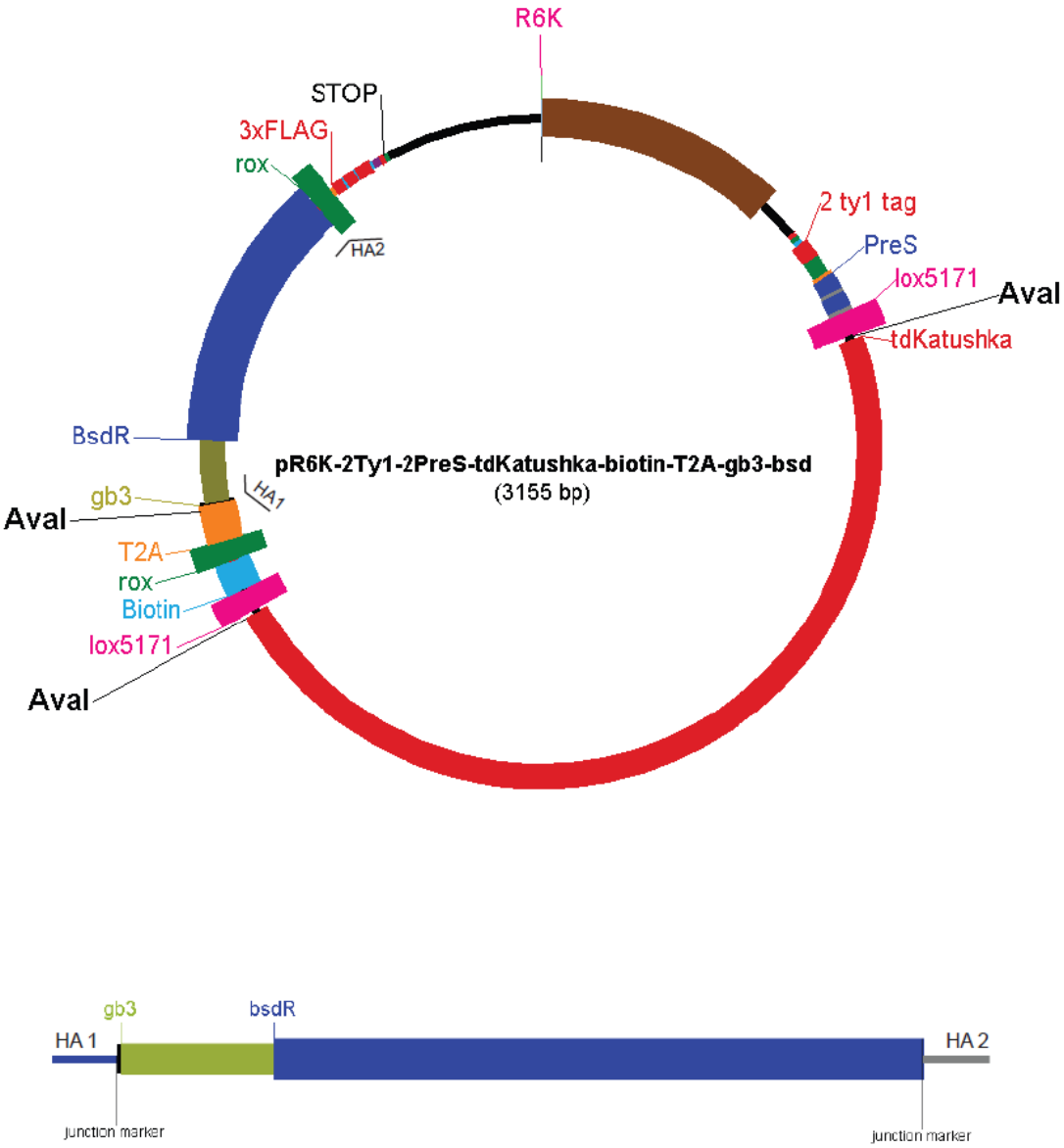

Figure 2.4. The pR6K-2Ty1-2PreS-tdKatushka-biotin-T2A-gb3-Bsd plasmid (top) was generated as a protein tagging construct in the Stewart lab, however it was convenient to generate a gb3-Bsd cassette. The plasmid was digested with AvaI and used a PCR template in a reaction with primers BSD1 and 2 to generate the gb3-Bsd PCR cassette **(bottom)**.

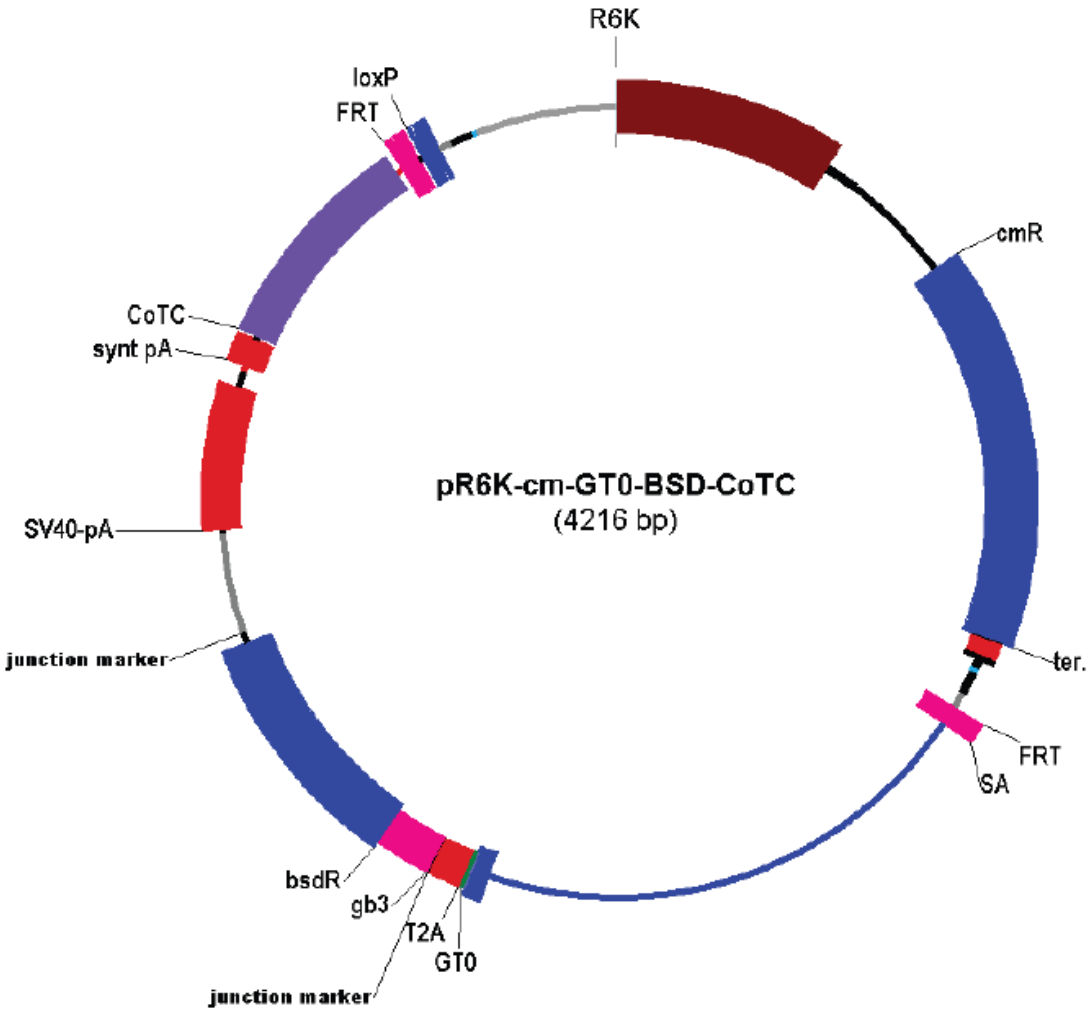

Figure 2.5. The pR6K-cm-GT0-Bsd-CoTC plasmid is the recombination product of the gb3-Bsd PCR product into the pR6K-cm-GT0-lacZneo-CoTC. R6K: origin of replication, cmR: chloramphenicol resistance gene, FRT: site recognised by the Flp recombinase, SA: splice acceptor, GT0: adapter sequence to maintain the T2A in-frame, T2A: sequence that instructs the ribosome to cleave the nascent peptide and begin synthesizing a new one, SV40-pA: poly-A signal from the SV40 virus, synt-Pa: poly-A signal, CoTC: transcription termination sequence, loxP: site recognised by the Cre recombinase. Junction markers flank the cassette that was introduced into this plasmid.

2.19.3. Modification of the *Mll2*-BAC

Initially, the BAC was modified by using an rpsL-genta cassette, obtained by PCR on the pR6K-photo-rpsL-genta plasmid (*Figure 2.6*). Primers *Mll2*-genta1 and 2 were used to amplify the rpsL-genta portion of the plasmid. These primers contain the HA1 and HA2 regions that will mediate the final cassette exchange and homology to intron1 of *Mll2* that will mediate recombineering into the BAC. Moreover, primer *Mll2*-genta1 contains a loxP site that will be situated 5' of the final cassette. The 3' loxP site was already on the pR6K-cm-GT0-lacZneo-CoTC as shown in *Figures 2.3* and *2.4*. The BAC host was (GB05-Red *E.coli*) cultured overnight at 37^{o}C with shaking and prepared for transformation/recombineering as described in paragraph 2.18.2 for the EC1000-pir116+pSC101-ETgA-tet. The BAC host was transformed with 200 ng of the rpsL-genta cassette by electroporation and cultured overnight in LB-agar plates containing 2 μg/ml gentamicin (Sigma). The introduction of the wild type rpsL restores sensitivity to streptomycin, allowing for counter-selection when the rpsL-genta cassette is exchanged for the gb3-Bsd. The resulting BACs were tested by single digestions with EcoRI, PvuII, BamHI and BglII. One clone was selected to proceed further. The BAC host was prepared again for transformation/recombineering and transformed by electroporation with 1μg of EcoRI/SwaI digested pR6K-cm-GT0-BSD-CoTC. The transformed bacteria were cultured on 200 μg/ml streptomycin and 30 μg/ml blasticidin S LB-agar plates. Single colonies were picked and tested by restriction digestion and sequencing. Unfortunately,

there were no correct clones. Due to unforeseen recombineering events the 3' of the cassette integrated in different regions of the BAC, possibly at the gb3-neo downstream of the *Mll2* gene. The structure of the final construct is shown in *Figure 2.7*.

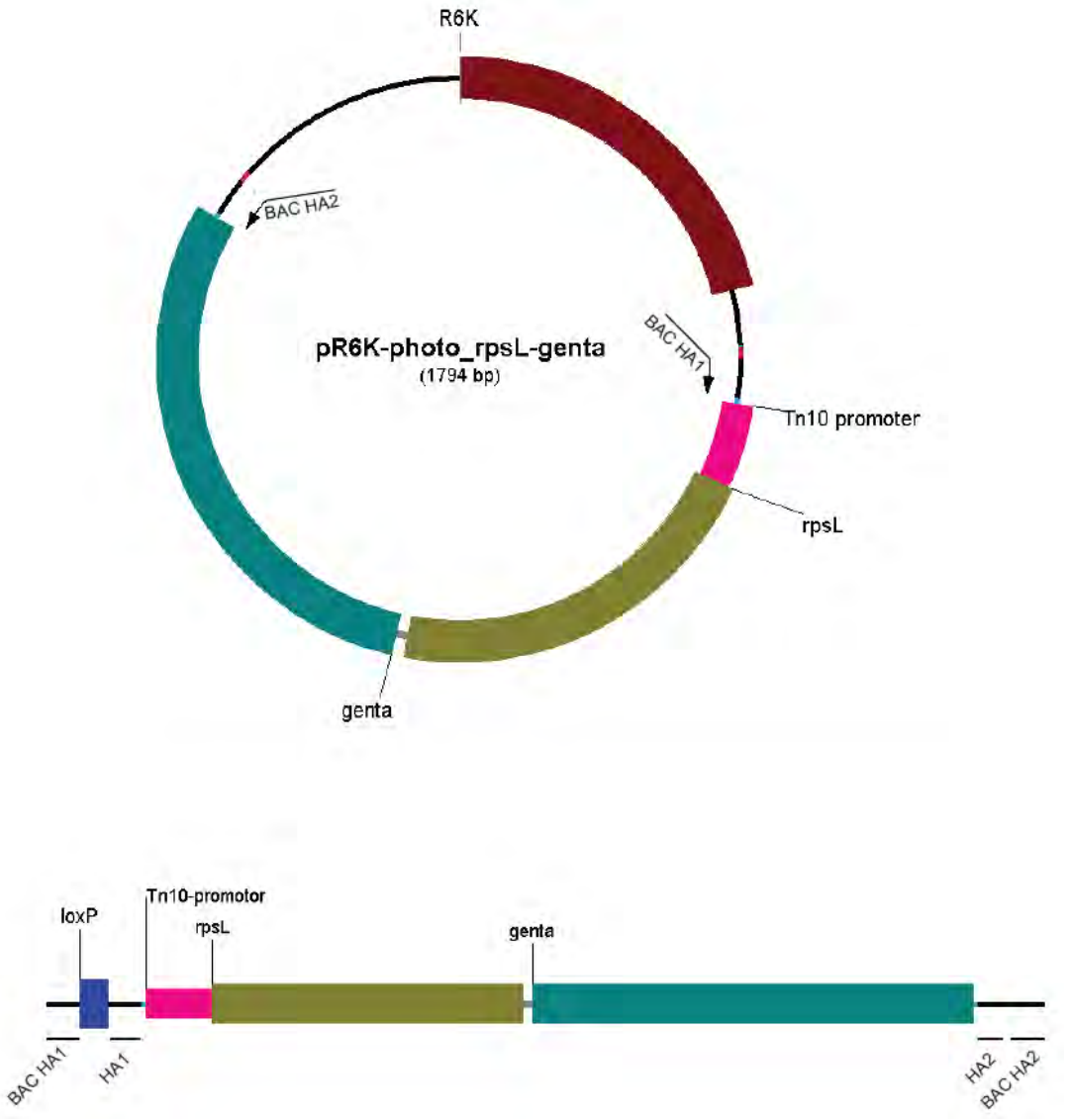

Figure 2.6. The pR6K-photo-rpsL-genta plasmid (top) is an adaptor cassette for recombineering allowing for selection of the recombined clones and counter selection when this cassette is swapped for the final cassette. The adaptor cassette was obtained by a PCR reaction on this plasmid using primers *Mll2*-genta1 and 2 **(bottom)**. Primer *Mll2*-genta1 contains a region homologous to the *Mll2* BAC (BAC HA1), a loxP site, the HA1 region and the primer region that is complementary to the 5' of the Tn-10 promoter, while primer *Mll2*-genta2 contains a BAC homology region (BAC HA2) the HA2 region and the primer sequence that is complementary to the 3' of genta. rpsL: *E. coli* wild type rpsL confers sensitivity to streptomycin, genta: gentamicin resistance gene.

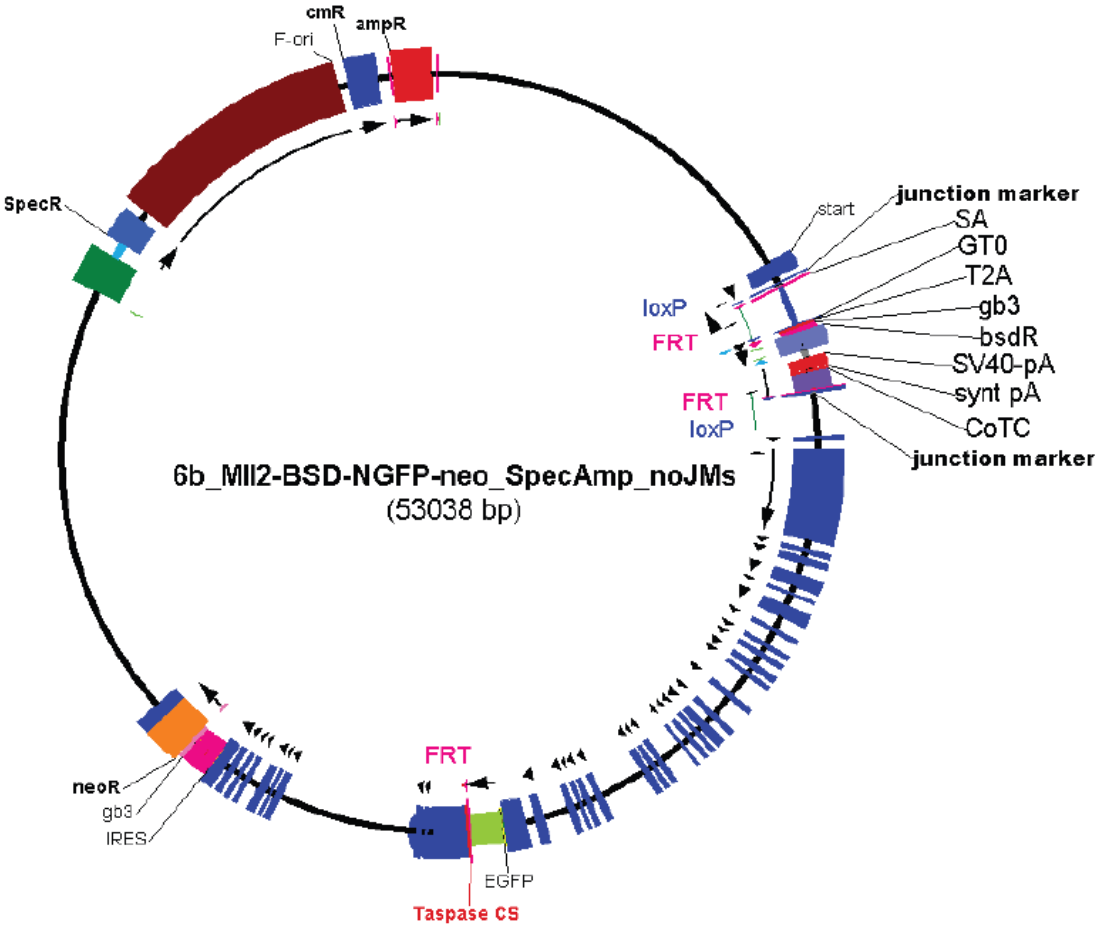

Figure 2.7. The final Mll2-Bsd-NGFP-neo-spec-amp construct would contain the stop cassette in intron1 trapping the transcript in the cassette and generating truncated mRNA. The cassette would be excisable in the *Mll2*$^{F/F}$ and *Mll2*$^{F/+}$ ES cells that express the Cre-ERT2, therefore allowing the full *Mll2* transcript to be generated from the BAC construct.

3. RESULTS

3.1. *Magoh2* transcriptional regulation and chromatin structure

The main aim of this work was to investigate the molecular details of how MLL2 is involved in driving transcription from the *Magoh2* CpG island promoter. To this end, we studied the chromatin structure and expression of this gene in wild type and *Mll2*$^{-/-}$ ES cells. The latter harbour a gene-trapping cassette in intron 1 of both alleles of *Mll2* (*Figure 3.1*). The nascent *Mll2* mRNA is spliced to the cassette which contains a neomycin resistance gene, two poly-adenylation signals and a transcriptional terminator, thus preventing expression of wild-type *Mll2.*

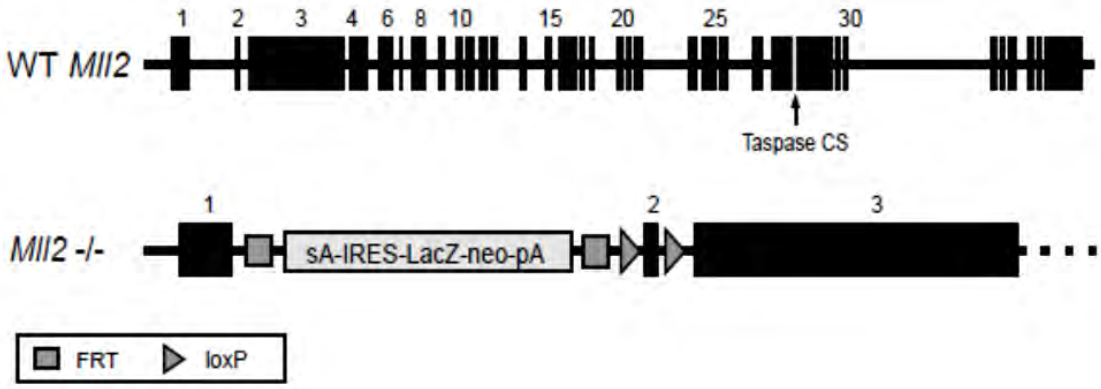

Figure 3.1. Schematic of the *Mll2* alleles in *Mll2*$^{-/-}$ cells. The mouse embryonic stem cell lines used harbour either the wild type *Mll2* allele (E14) or the *Mll2*$^{-}$ allele. The *Mll2*$^{-}$ allele contains a STOP cassette within intron1 of *Mll2*. The STOP cassette is composed of a splice acceptor (sA), an internal ribosome entry sequence (IRES), the β-galactosidase gene (*lacZ*), a neomycin resistance gene (*neo*), two poly-adenylation signals (pA) and a transcriptional terminator sequence (*not shown*). Black boxes indicate the *Mll2* exons. MLL2 is post-translationally cleaved by the Taspase 1 protease. The Taspase cleavage site is indicated by the arrow.

3.1.1. *Magoh2* is transcriptionally silenced in the absence of MLL2

To test whether *Magoh2* expression depended on the presence of MLL2, we measured *Magoh2* steady-state mRNA levels by reverse-transcription (RT) and quantitative real-time PCR (QPCR) in E14 wild type and *Mll2*$^{-/-}$ ES cells. Our results demonstrate that *Magoh2* expression was reduced to near background levels in the *Mll2*$^{-/-}$ ES cells (detected at C_T >30). This is consistent with previously reported findings (Glaser et al., 2009) suggesting that MLL2 is directly or indirectly required for *Magoh2* expression (*Figure 3.2*).

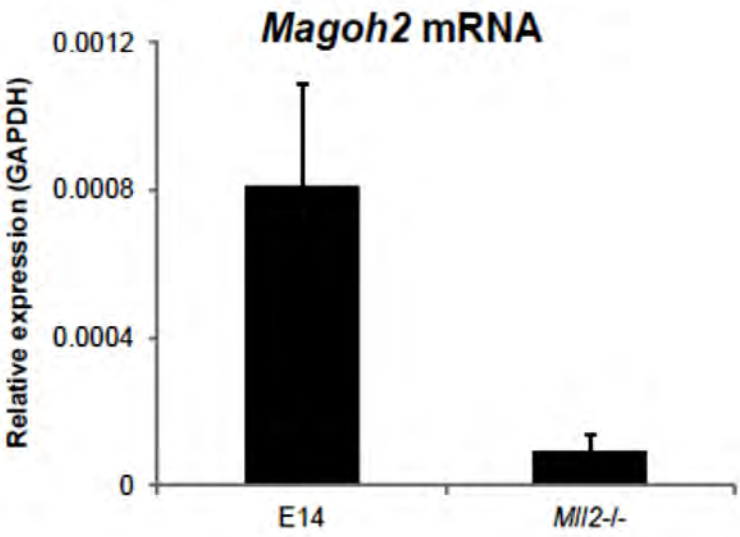

Figure 3.2. *Magoh2* is transcriptionally silenced in *Mll2*$^{-/-}$ ES cells. Reverse transcription and quantitative PCR with primers specific for *Magoh2.* Bars represent the mean +sd of 6 measurements (3 biological x 2 technical replicates)

3.1.2. DNaseI hypersensitive sites on the *Magoh2* locus

To investigate the reason for *Magoh2* silencing, we first had to identify potential *cis* regulatory elements regulating *Magoh2* expression. To this end we employed DNaseI hypersensitive site (DHS) mapping. Three conserved intronic regions were identified (through the UCSC genome browser's mammalian conservation plots) which could play a role in *Magoh2* transcriptional regulation and were termed conserved sequence (CS) 1, 2 and 3 (*Figure 3.3 Top panel*). The assay was designed to encompass the entire *Magoh2* gene and ~7Kb upstream (up to the end of the neighbouring *Styk* gene) within one BamHI restriction fragment. The region downstream of the *Magoh2* gene was not examined as it is a 30Kb long gene desert that is enriched in repeat elements, making analyses technically challenging. Hybridization with probes on either side of the BamHI restriction fragment revealed the presence of two DNaseI hypersensitive sites, termed HSp and HS1 (*Figure 3.3 Middle and Lower panel*).

HSp forms over the *Magoh2* promoter only in E14 wild type cells. The absence of HSp in the *Mll2*$^{-/-}$ cells suggests that MLL2 is involved in the maintenance of the "open chromatin" conformation of the *Magoh2* promoter. HS1 was found to be in close proximity to conserved sequence 1 (CS1). To accurately map the position of HS1 we designed a higher resolution DHS mapping experiment, using probes closer to HS1 (*Figure 3.4*). Southern blotting revealed that the HS1 does not overlap CS1; rather

it forms over a simple repeat element. The increased DNaseI sensitivity of this site may be because the DNA sequence of the repeat element is unfavourable for nucleosome assembly. If that were the case we would not expect it to change upon MLL2 depletion. Although CS1, 2 and 3 contain putative binding sites for transcription factors, they do not seem to be active in ES cells. These elements may act as enhancers either for *Magoh2* or other nearby genes in other cell-types.

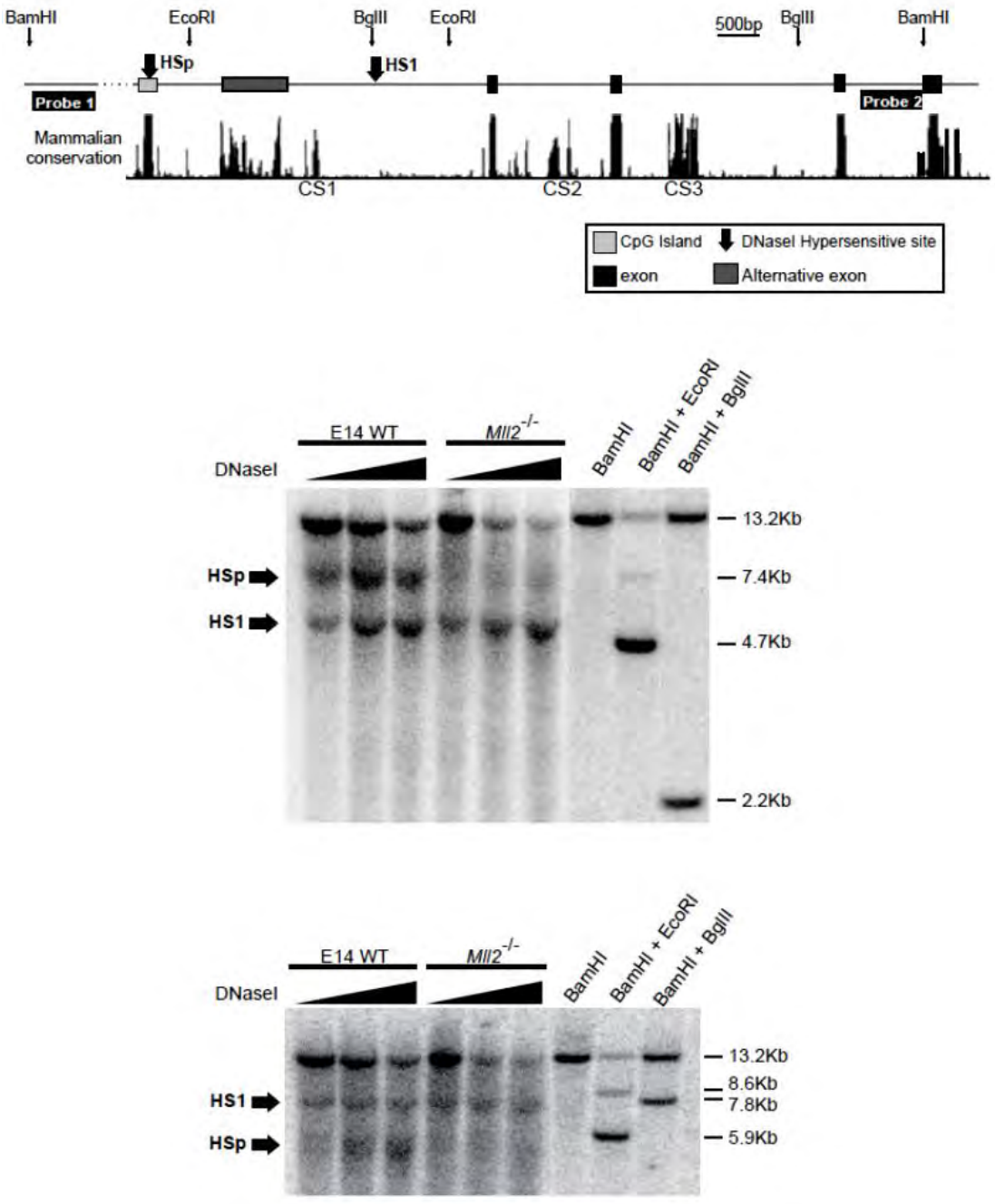

Figure 3.3. DNaseI hypersensitive sites on the *Magoh2* locus in *Mll2*$^{-/-}$ and E14 wild type ES cells. Top panel: Schematic of the *Magoh2* locus. CS1, 2 and 3 mark three conserved regions within the locus that could potentially act as *cis* regulatory elements. Probe1 and 2 mark the two probes used for the Southern blots shown in the lower and middle panels respectively. Arrows show the positions of the two DNaseI hypersensitive sites detected. The last three lanes serve as a DNaseI non-digested control (BamHI) and as molecular weight markers (BamHI+EcoRI/BglII). **Middle panel:** Hybridization with Probe 2. Arrows show the two DNaseI hypersensitive sites detected. **Lower panel:** Hybridization with Probe 1 verifies the presence of the two DNaseI hypersensitive sites.

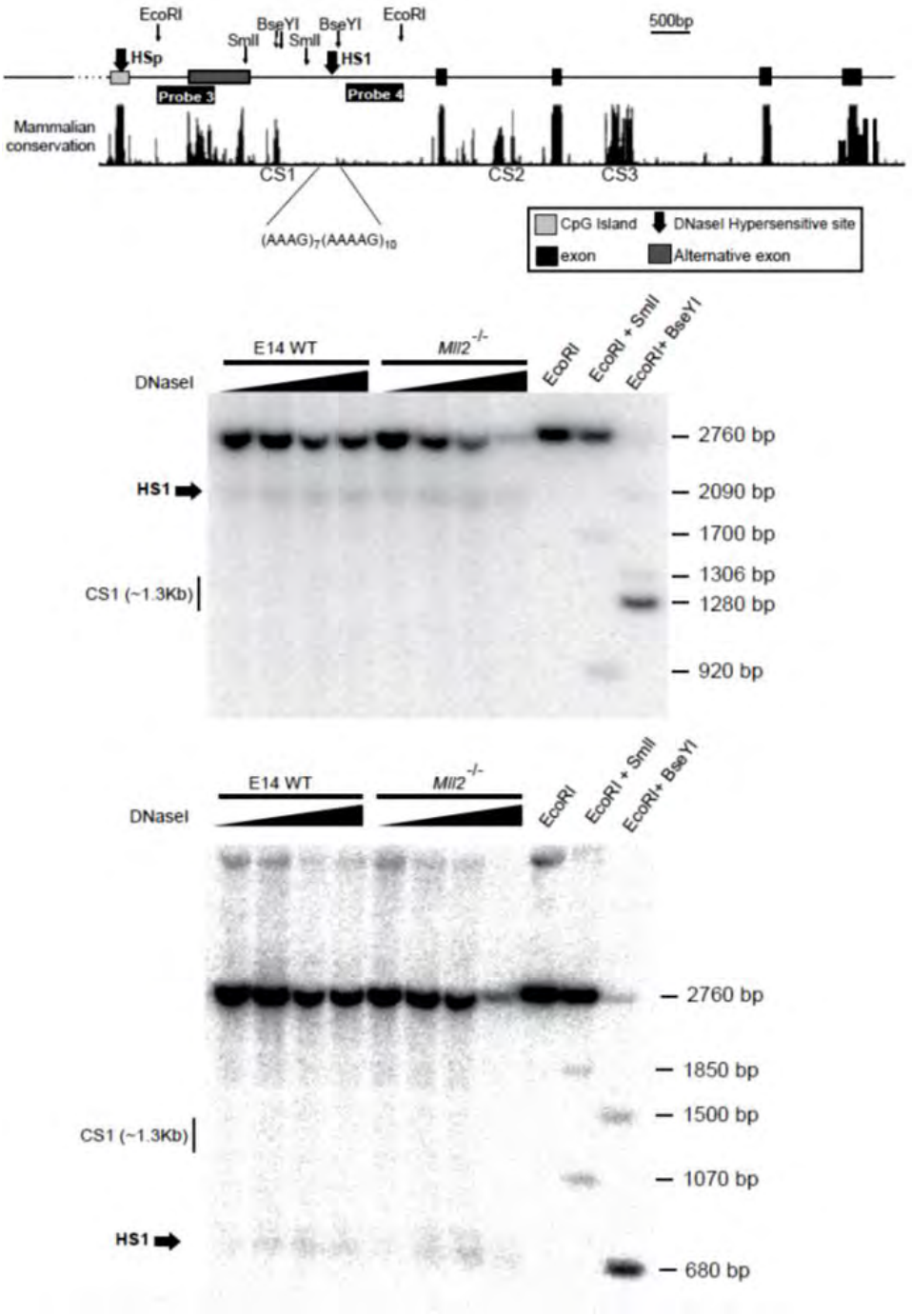

Figure 3.4. HS1 does not overlap CS1. Top panel: Schematic of the *Magoh2* locus. Probe 3 and Probe 4 mark the two probes used for the Southern blots shown in the middle and lower panels respectively. Arrows show the positions of the two DNaseI hypersensitive sites detected. The repeat element is illustrated in brackets. The last three lanes serve as a DNaseI non-digested control (EcoRI) and as molecular weight markers (EcoRI+SmlI/BseYI). **Middle panel:** Hybridization with Probe 3. HS1 is indicated by the arrow. The position of CS1 is illustrated by the bar. **Lower panel:** Hybridization with Probe 4.

3.1.3. The *Magoh2* promoter harbours binding sites for SP and CREB family transcription factors

To elucidate how the *Magoh2* promoter is activated we had to identify which sequence specific transcription factors bind this promoter. Potential transcription factor binding sites were identified by *in silico* sequence analyses. The prediction algorithms are publicly available through the Pasteur Institute's website (TFScan, bioweb.pasteur.fr). The transcription factors that could potentially bind the *Magoh2* promoter are illustrated in *Figure 3.5.*

Figure 3.5. Primary DNA sequence of the *Magoh2* CpG island promoter. Capital letters mark the CpG island. The CpG island is well defined as it is flanked on the 5' by a series of AC repeats and a stretch of T on the 3'. This figure also illustrates *in silico* predicted transcription factor binding sites (TFScan, www.bioweb.pasteur.fr). The TSS shown is as it is annotated in the UCSC genome browser and overlaps a CAGE peak as shown by data produced by the FANTOM consortium.

To investigate whether the *in silico* predicted TFs could bind the *Magoh2* promoter *in vitro* we performed electrophoretic mobility shift assays. Overlapping double stranded DNA probes were designed to cover the *Magoh2* promoter and were incubated with nuclear protein extracts from E14 wild type ES cells (*Figure 3.6*). Protein binding was detected for probes 1 and 7 that contain SP1/3 and CREB binding sites respectively. Weak protein binding was observed for probes 5, 8 and 9. In all cases, the observed protein binding could be disrupted by use of a 100-fold excess of unlabelled probe.

To verify that SP family members are the proteins binding to probe 1, the assay was repeated. This time we used antibodies specifically targeting either SP1 or SP3 to generate a super-shift (*Figure 3.7*). A probe containing the SP consensus sequence was used as a competitor (lane 3) and could completely abolish protein binding on probe 1, suggesting that protein binding is localised on the predicted SP sites. The super-shifts observed when α-SP antibodies were added to the binding reactions (lanes 4 and 5) verify that SP1 and SP3 indeed bind to the *Magoh2* promoter *in vitro*.

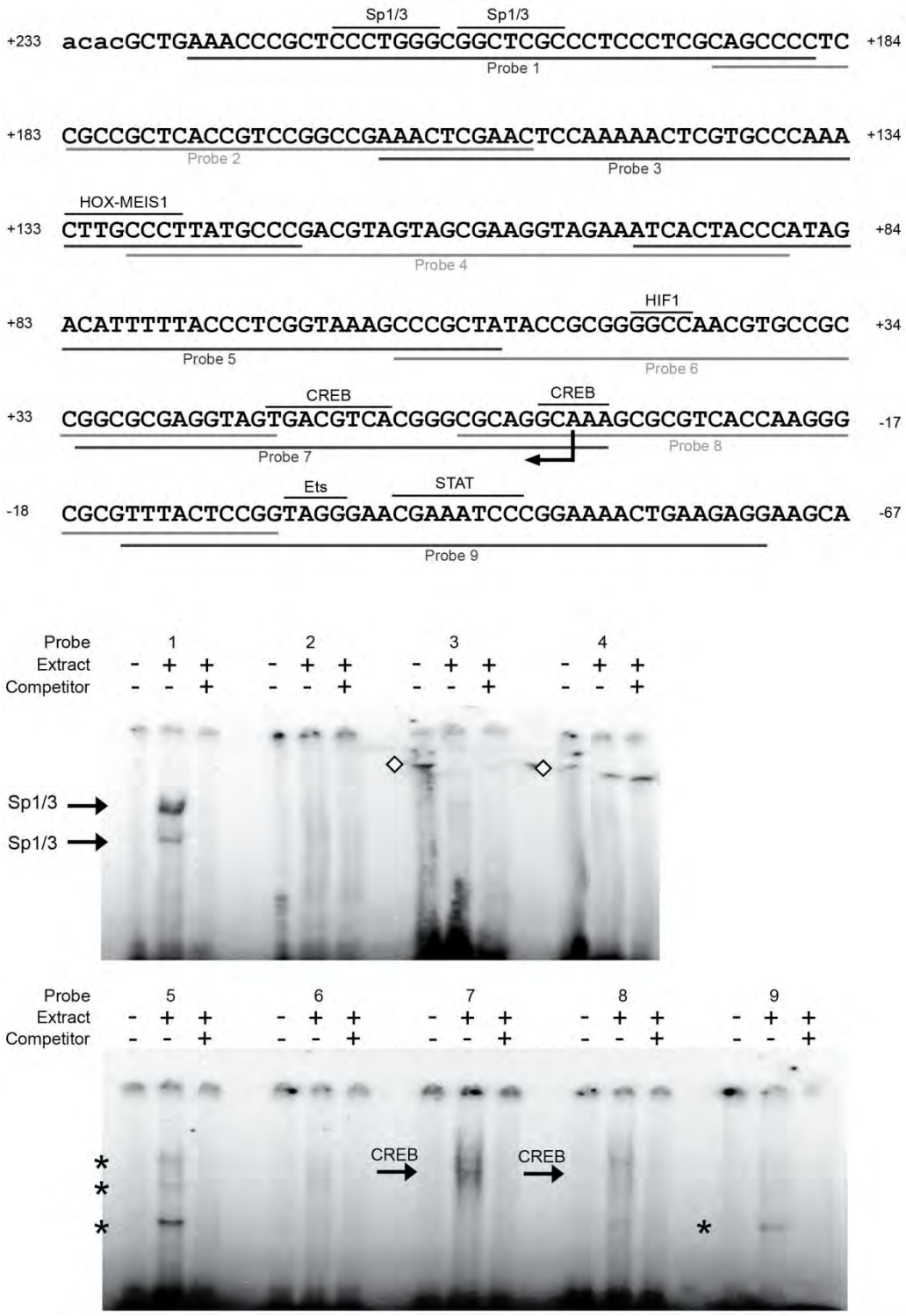

Figure 3.6. SP1/3 and CREB bind the *Magoh2* promoter *in vitro.* Top panel: Double stranded DNA probes (1-9) were designed to cover the entire *Magoh2* CpG island promoter. **Lower panel:** Electrophoretic mobility shift assays using the probes shown above. *In vitro* protein binding to probe1 and probe7 is shown by arrows. Weak protein binding was observed on probes 5 and 9 indicated by the asterisks. Diamonds (probes 3 and 4) indicate non-specific signal as a result of mechanical damage to the gel. All the probes were titrated (1-15 pmoles) so that 20,000cpm/lane were loaded on the gel. The non-labelled competitor was used at 100-fold excess of the respective labelled probe.

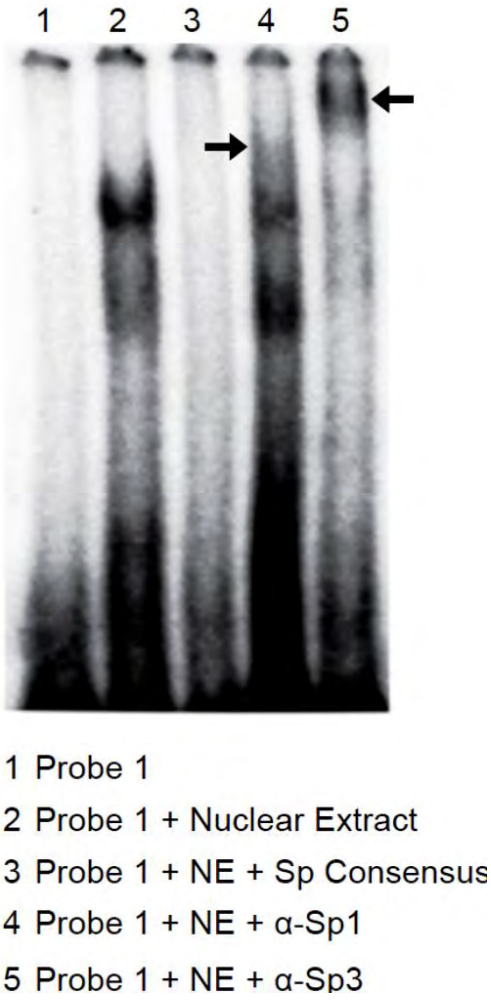

Figure 3.7. SP1 and SP3 are verified as being able to bind the *Magoh2* promoter. The electrophoretic mobility shift assay was repeated for probe1. This time a probe containing the SP1/3 consensus sequence (courtesy of Dr. S. Bowers) was used as non-labelled competitor in a 100-fold excess of the labelled probe (lane 3). Antibodies against SP1 and SP3 (1µg each) were added to the binding reaction (lanes 4 and 5 respectively). The position of the super-shifted bands is indicated by arrows. The probe was titrated such that 20,000cpm/lane were loaded on the gel.

3.1.4. The SP and CREB sites are not occupied *in vivo*

The *in vivo* presence of SP and CREB family members on the *Magoh2* promoter was investigated by *in vivo* dimethyl-sulphate (DMS) footprinting. DMS is a highly reactive molecule that methylates guanine residues on DNA at the N7 position (Brookes and Lawley, 1963). Subsequent alkaloid and heat treatment cleaves DNA at these modified sites (Maxam and Gilbert, 1977). G-residues within TF binding sites will be protected from DMS mediated methylation if the TF is bound *in vivo* and thus escape alkaloid/heat induced cleavage. Furthermore, TF binding can lead to increased methylation of G residues immediately adjacent to the TF binding site which can then be detected as increased band intensity. DNA fragments are isolated, amplified, radiolabelled and separated on high resolution denaturing acrylamide gels. When comparing *in vivo* to *in vitro* DMS treated DNA, loss of specific bands in the *in vivo* sample indicates that this particular G residue is bound by a protein and is not accessible to DMS. The advantage of this technique is that it allows us to investigate binding of essentially any factor to a specific region within a single assay. These experiments did not reveal any protection from DMS-mediated methylation of G residues within the SP1/3 or CREB binding sites (*Figure 3.8*), suggesting that if these factors bind the *Magoh2* promoter *in vivo,* the binding is weak and/or transient.

Moreover, little difference could be reproducibly observed between the E14 wild type and *Mll2*$^{-/-}$ ES cells over the entire *Magoh2* promoter. Only one G residue appeared to be consistently (over 3 experiments) protected from DMS-mediated methylation (+122bp) in the wild type cells when compared to G-reaction or Mll2$^{-/-}$ cells. Although this protection was reproduced in three independent experiments, no known TF was predicted to bind within that region.

To further investigate the role of SP1/3 in *Magoh2* regulation, *Magoh2* steady state mRNA levels were measured in *Sp1*$^{-/-}$, rescued *Sp1*$^{-/-}$ (*Sp1R*) and *Sp3*$^{-/-}$ ES cells (*Figure 3.9*). *Magoh2* mRNA levels were unaltered in the *Sp3*$^{-/-}$ ES cells when compared to wild type E14 cells. The *Sp1*$^{-/-}$ cells exhibited a marked overexpression of *Magoh2*, but the same overexpression was observed in the *Sp1R* cells. Since re-expression of *Sp1* did not attenuate the expression levels of *Magoh2*, it is more likely that this overexpression is due to clonal differences between the individual cell lines rather than a specific effect of SP1 depletion. We conclude that SP1 and SP3 most likely do not have individual non-redundant roles in *Magoh2* regulation.

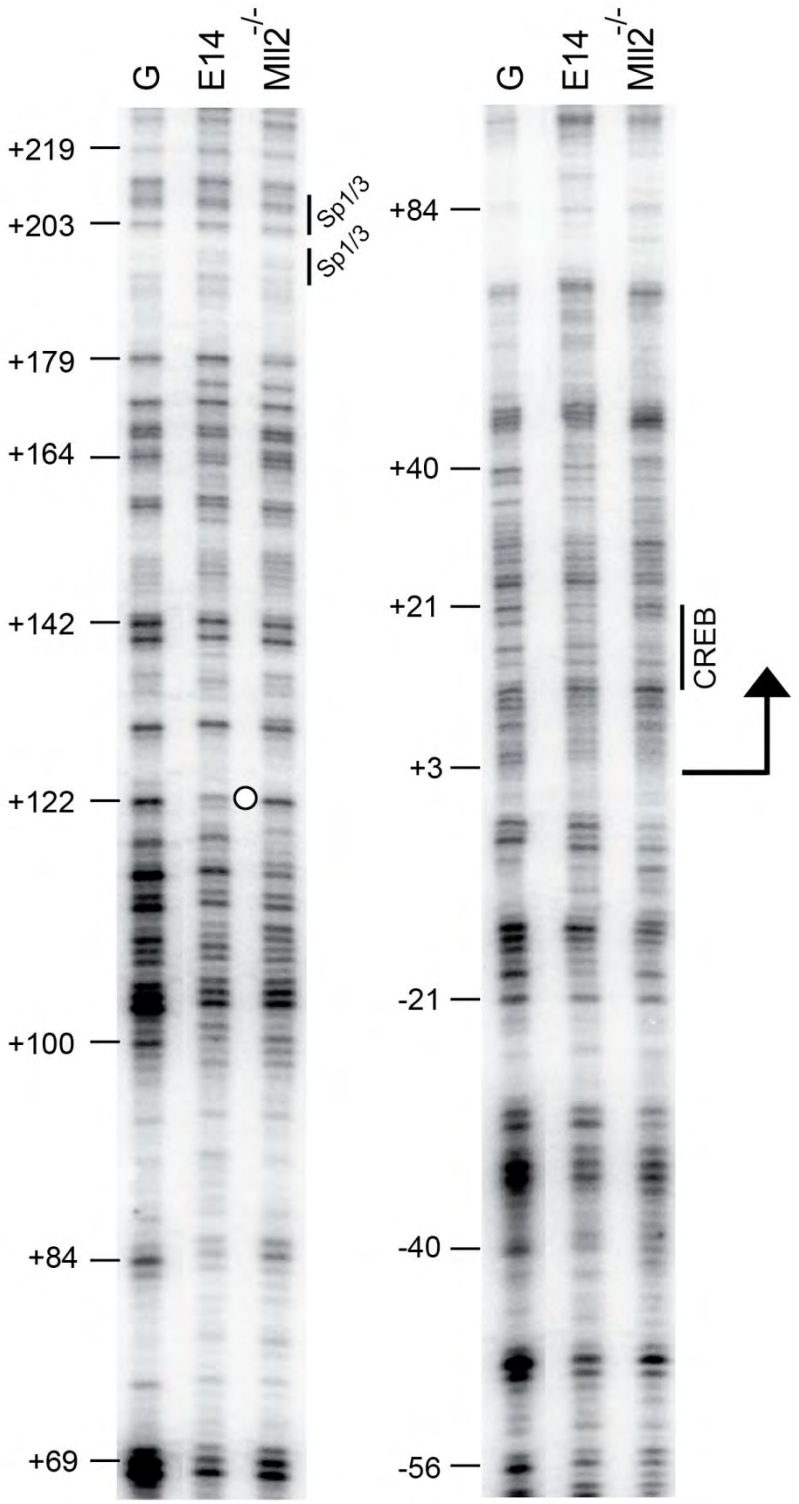

Figure 3.8. SP1/3 and CREB do not bind the *Magoh2* promoter *in vivo.* Dimethyl-sulphate footprinting over the *Magoh2* promoter in E14 wild type ES cells and the *Mll2*$^{-/-}$ ES cells. Annotation is relative to the transcription start site, indicated by the arrow. G: G-reaction (*in vitro* DMS treated genomic DNA). The CREB and the two SP1/3 binding site are shown by the bars. White circle indicates a G nucleotide that was consistently protected in 3 individual experiments.

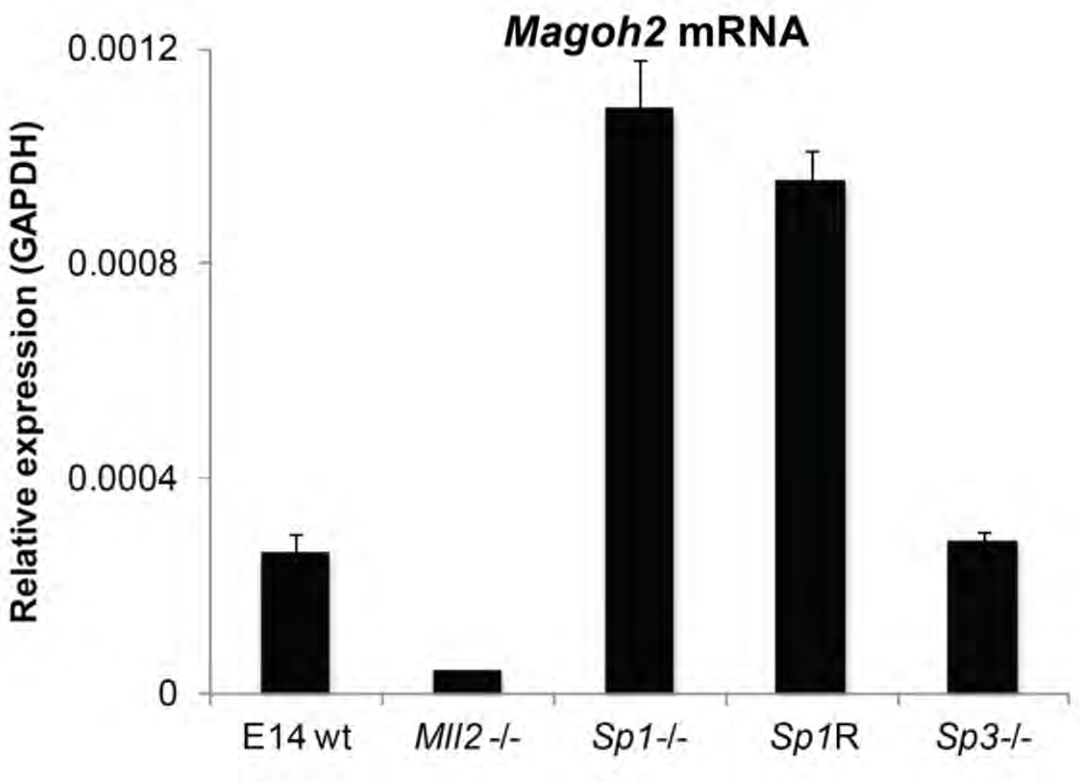

Figure 3.9. SP1 and SP3 do not appear to have an individual functional role in *Magoh2* regulation. *Sp1* and *Sp3* knock-out ES cell lines were used to measure *Magoh2* expression by reverse transcription and quantitative PCR. *Sp1R* are *Sp1*$^{-/-}$ ES cells rescued with a *Sp1* expressing plasmid. The bars represent the mean +sd of 4 measurements (2 biological x 2 technical replicates).

3.1.5. The *Magoh2* transcription start site is occupied in the presence of MLL2

To identify other proteins that may bind the *Magoh2* promoter we employed DNaseI footprinting (*Figure 3.10*). The methodology of this assay is very similar to DMS footprinting, however as DNaseI is much larger than DMS, it allows detection of DNA regions that are covered by a protein complex and gives a measure of overall chromatin accessibility, as indicated by the intensity of the DNaseI produced cleavage products (Galas and Schmitz, 1978). The results demonstrated that a protein or more likely a protein complex is tethered to the TSS region, providing protection from DNaseI digestion, as compared to an *in vitro* DNaseI digested control. This protection spans a 37bp stretch and is flanked by DNaseI hyper-sensitive regions on both sides. It is likely that this DNaseI protection is due to the presence of a RNA polymerase II pre-initiation complex on the *Magoh2* promoter. Cells lacking MLL2 exhibit no such binding, suggesting that MLL2 is required for recruitment of these protein factors to the *Magoh2* promoter. Additionally, the SP and CREB binding sequences on the *Magoh2* promoter were found to be accessible to DNaseI (*data not shown*), further suggesting that these factors are not binding the *Magoh2* promoter *in vivo*.

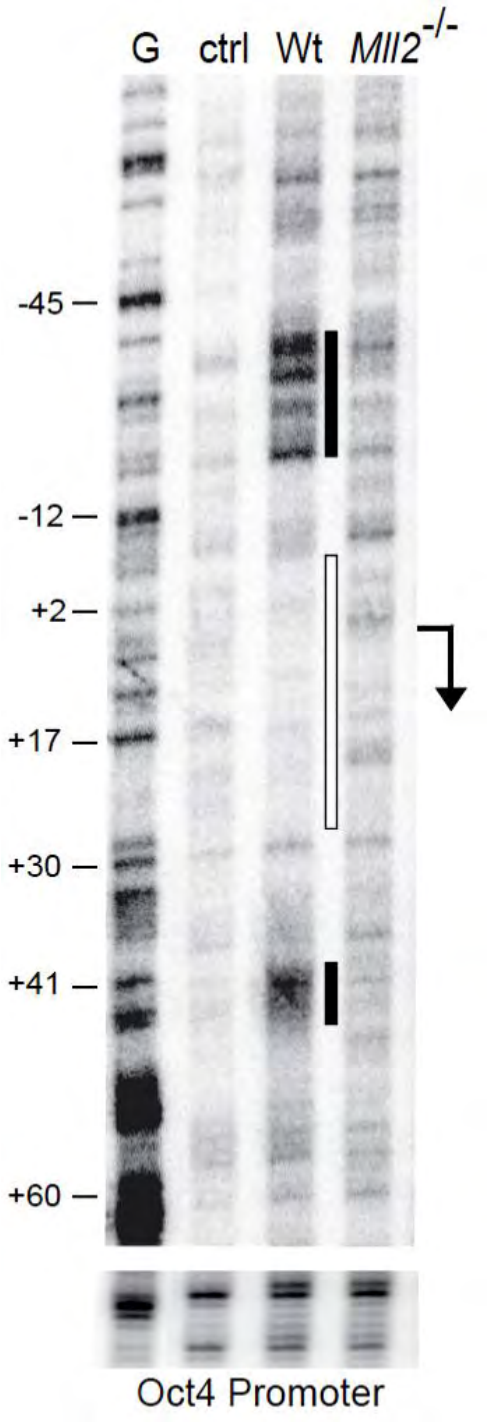

Figure 3.10. The *Magoh2* transcription start site is occupied by a large protein complex. DNaseI footprinting of the *Magoh2* CpG island promoter. The region protected from DNaseI digestion spans from -8bp to +29bp, as indicated by the white bar. The protected region is flanked by hypersensitive regions (black bars) G: G-reaction, ctrl: *in vitro* DNaseI digested DNA. Arrow: transcription start site. Annotation is relative to the transcription start site. Primers on the *Oct4* promoter were used to verify equal DNaseI digestion of the samples.

3.1.6. Loss of active histone marks and RNA Pol II in *Mll2*$^{-/-}$ cells

As histone modifications have been proposed to have some role in transcriptional regulation, we wanted to investigate whether *Magoh2* silencing was accompanied by changes in the chromatin landscape, as suggested for other genes after MLL1 depletion. To this end we employed chromatin immunoprecipitation studies to examine the chromatin state of the *Magoh2* promoter (*Figure 3.11*). These assays revealed that when *Magoh2* is expressed, nucleosomes on the *Magoh2* promoter bear the H3K4me$_3$ mark. The levels of this modification were greatly reduced in the absence of MLL2, consistent with its H3K4 methyltransferase activity. As H3K4me$_3$ is considered to be a histone mark of active transcription, this finding is consistent with the observed *Magoh2* transcriptional silencing in *Mll2*$^{-/-}$ ES cells presented in *Figure 3.2*. Additionally, acetylation of histone H3 on lysine 9 (H3K9ac) – another mark associated with active transcription – was also absent from the *Magoh2* promoter in *Mll2*$^{-/-}$ cells. Immunoprecipitation with an antibody against RNA polymerase II demonstrated that the transcriptional machinery engages with the *Magoh2* promoter only in the presence of MLL2, further supporting the DNaseI footprinting experiments presented in *Figure 3.10*. These experiments are consistent with the hypothesis that MLL2 is required to establish the active chromatin environment on the *Magoh2* promoter. Additionally, the data suggest that MLL2 recruitment and H3K4me$_3$ deposition coincide with recruitment of the basal transcription machinery to the *Magoh2* promoter.

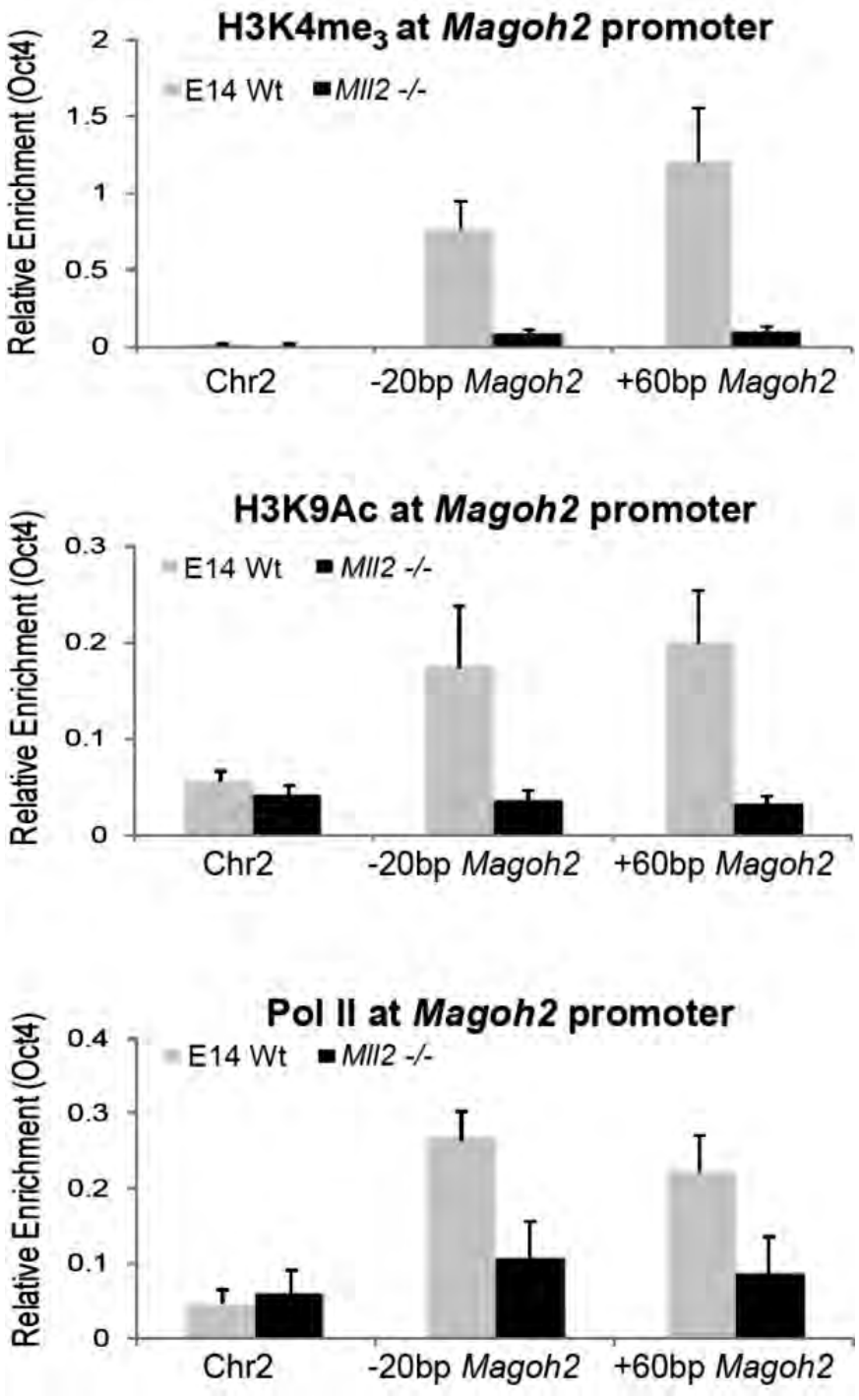

Figure 3.11. Loss of active chromatin marks and RNA polymerase II on the *Magoh2* promoter in the absence of MLL2. ChIP assays were performed using antibodies against $H3K4me_3$ **(top)**, H3K9ac **(middle)** and the RNA polymerase II CTD **(lower)**. The immunoprecipitated DNA was amplified by QPCR with a primer set on an inactive region on chromosome 2 (negative control) and two primer sets on the *Magoh2* promoter centred around -20bp and +60bp as indicated. Data for histone modifications are corrected for nucleosome density, as determined by a histone H3 immunoprecipitation. Bars represent the mean +sd of biological triplicates. A value of 1 is approximately 19% of Input.

3.1.7. Acquisition of DNA methylation on the *Magoh2* CpG island in the absence of MLL2

Extensive research has shown that when CpG islands become transcriptionally silenced they do not necessarily acquire DNA methylation. With regards to the *Magoh2* CpG island promoter, Glaser et al. (2009) – using an inducible *Mll2* knock-out system – reported that this promoter is methylated 12 days after *Mll2* deletion. Our initial experiments using the methylation sensitive restriction enzyme HpaII and PCR over a single HpaII site suggested that this is indeed the case (*data not shown*). To further validate this result, total genomic DNA from E14 wild type and *Mll2*$^{-/-}$ cells was treated with sodium bisulphite, amplified by PCR and sequenced on a pyrosequencer (*Figure 3.12*). The result clearly shows an increase in cytosine modification levels in the *Mll2*$^{-/-}$ cells.

A significant amount of cytosine modification was detected in the wild type cells as well. However, the method employed cannot distinguish between methylated and hydroxymethylated cytosines (Jin et al., 2010). It is likely that the modification responsible for this phenomenon in the E14 wild type cells is not methylation, as promoter methylation almost invariably leads to transcriptional silencing.

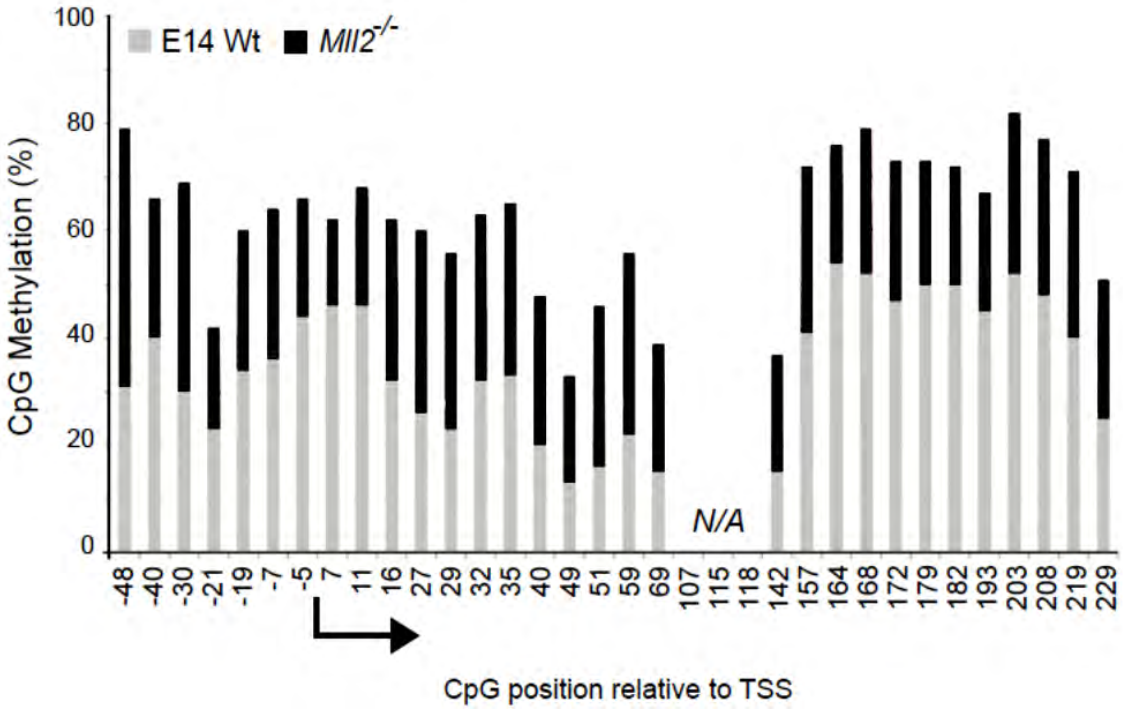

Figure 3.12. DNA methylation is increased on the *Magoh2* promoter in the absence of MLL2. Genomic DNA from E14 wild type and *Mll2*$^{-/-}$ was treated with sodium bisulphite, amplified with primers specific for the bisulphite converted *Magoh2* CpG island sent for pyrosequencing**.** The overlaid bars represent the percentage of modified cytosine residues within each sample. Each point on the X-axis represents an individual CpG dinucleotide. The *Magoh2* transcription start site is indicated by the arrow.

3.2. Chromatin dynamics at the *Magoh2* promoter

3.2.1. An inducible *Mll2* knock-out system

So far we showed that *Magoh2* expression, normal protein occupancy on the promoter and lack of DNA methylation depended on the presence of MLL2. To gain insight into the causal relationships of these observations we employed an inducible knock-out system. This system would allow us to perform kinetic experiments over a time-course to establish the order of events leading from the active open-chromatin state as seen in E14 WT cells to the silenced and DNA methylated state seen in $Mll2^{-/-}$ cells. The $Mll2^{F/F}$ and $Mll2^{F/+}$ ES cells ubiquitously express the Cre recombinase fused to a mutant oestrogen receptor ligand-binding domain (Cre-ERT2) from the *Rosa26* locus. Also, the wild type *Mll2* alleles have been modified to yield the $Mll2^{F}$ allele (*Figure 3.13*). In this allele, exon 2 of *Mll2* is flanked by loxP sites (the Cre cognate sequence). Activation of the Cre recombinase by addition of 4-hydroxytamoxifen (OHT) to the culture medium results in deletion of *Mll2* exon 2, causing a frame shift and introducing stop codons within exon 3. Consequently, no functional MLL2 can be synthesised in the cells after OHT induction. The $Mll2^{F/F}$ cells lose both functional *Mll2* alleles upon Cre activation, while the $Mll2^{F/+}$ lose only one allele and are used as a control cell line to exclude any OHT/Cre induced effects. The recombined (null) allele is hereafter termed $Mll2^{Fc}$.

The efficiency and kinetics of Cre-mediated recombination were examined by Southern blotting. As illustrated in *Figure 3.13,* recombination is complete 24 hours post OHT induction. Importantly, no outgrowth of non-recombined cells was observed up to 8 days post induction, as detected by the absence of the *Mll2*^{F} allele.

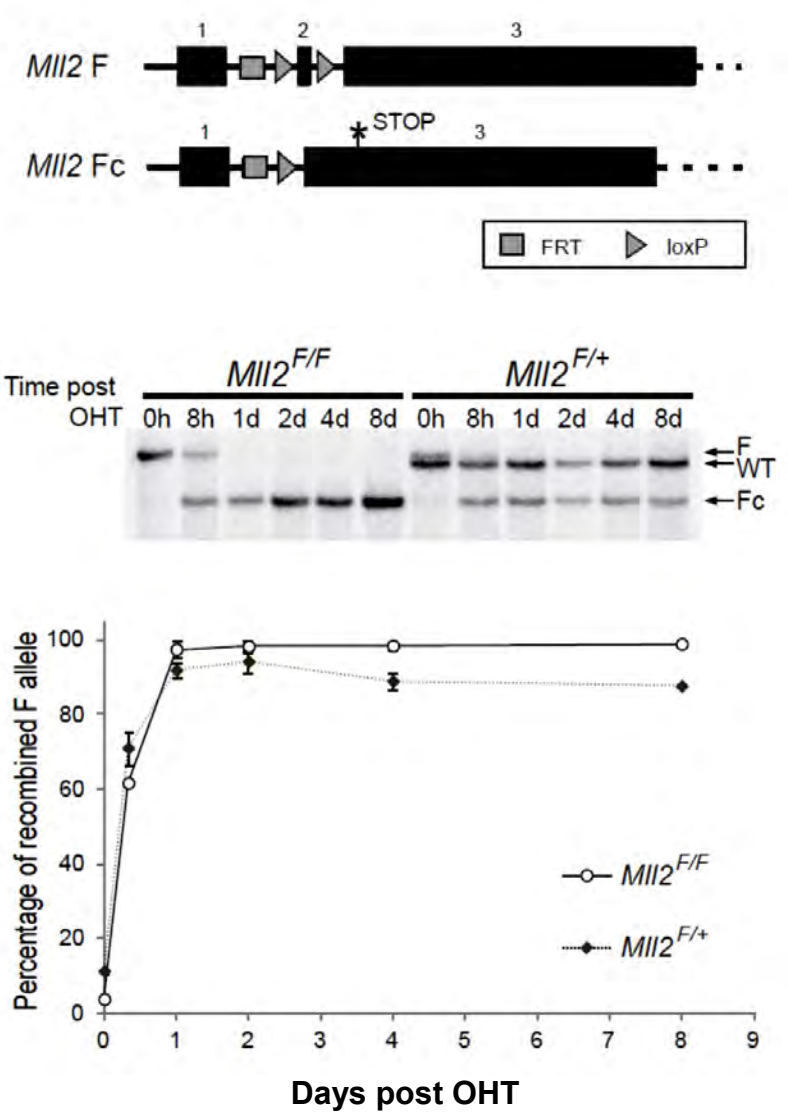

Figure 3.13. The inducible *Mll2* knockout system. Top panel: ES cells carrying an inducible *Mll2* knockout allele (*Mll2*^{F}) express the Cre recombinase fused to a mutant oestrogen receptor ligand-binding domain (Cre-ERT2) from the ubiquitous *Rosa26* locus. These cells have one or two conditional alleles (*Mll2*$^{F/+}$ and *Mll2*$^{F/F}$ respectively) of *Mll2*. The recombined allele is termed *Fc*. **Middle panel:** *Mll2* Southern blot after OHT treatment. The three different alleles for *Mll2* can be distinguished. *F*: conditional allele, WT: wild type allele, *Fc*: recombined null allele. **Lower panel:** Quantitation of the above Southern blot. The intensity quantification was performed using the QuantityOne software suite (Biorad). Error bars represent the standard deviation of 2 independent experiments.

3.2.2. Reduced proliferation of ES cells post *Mll2* deletion

Previous studies (Milne et al., 2005b; Takeda et al., 2006; Tyagi et al., 2007) have reported that MLL1 may be involved in cell cycle progression and regulation of proliferation. We hypothesised that the highly related MLL2 protein may have similar functions. To test this, we recorded cell numbers over a time-course, as it was the simplest way of measuring growth rates. This experiment demonstrated that *Mll2*$^{Fc/Fc}$ cells exhibit slower growth rates than their non-induced *Mll2*$^{F/F}$ counterparts (*Figure 3.14*), especially between 2 and 4 days post OHT induction. Importantly, loss of only one allele of *Mll2* did not seem to affect cell proliferation rates, as no differences were observed in the OHT induced (*Mll2*$^{Fc/+}$) versus non-induced *Mll2*$^{F/+}$ cells.

To find the reason for this apparent proliferation defect, we examined the cell cycle stage distribution of these cells to identify whether MLL2 plays a role in this process, as it has been suggested for MLL1. The cell cycle stage distribution was examined by FACS analysis of methanol fixed, propidium iodide stained *Mll2*$^{F/F}$ cells, over a 4-day time-course. The cells were separated into three populations based on their DNA content. The percentages of cells in each stage were plotted in a bar chart (*Figure 3.15*). This experiment revealed that there is no apparent block in the cell cycle after *Mll2* deletion. Further investigation revealed that cells lacking MLL2 exhibit slightly increased apoptosis rates (*Figure 3.16*), as measured by FACS analysis of Annexin V stained *Mll2*$^{F/F}$ ±OHT ES cells. This

increase in apoptosis could account for the lower cell counts shown in *Figure 3.14*.

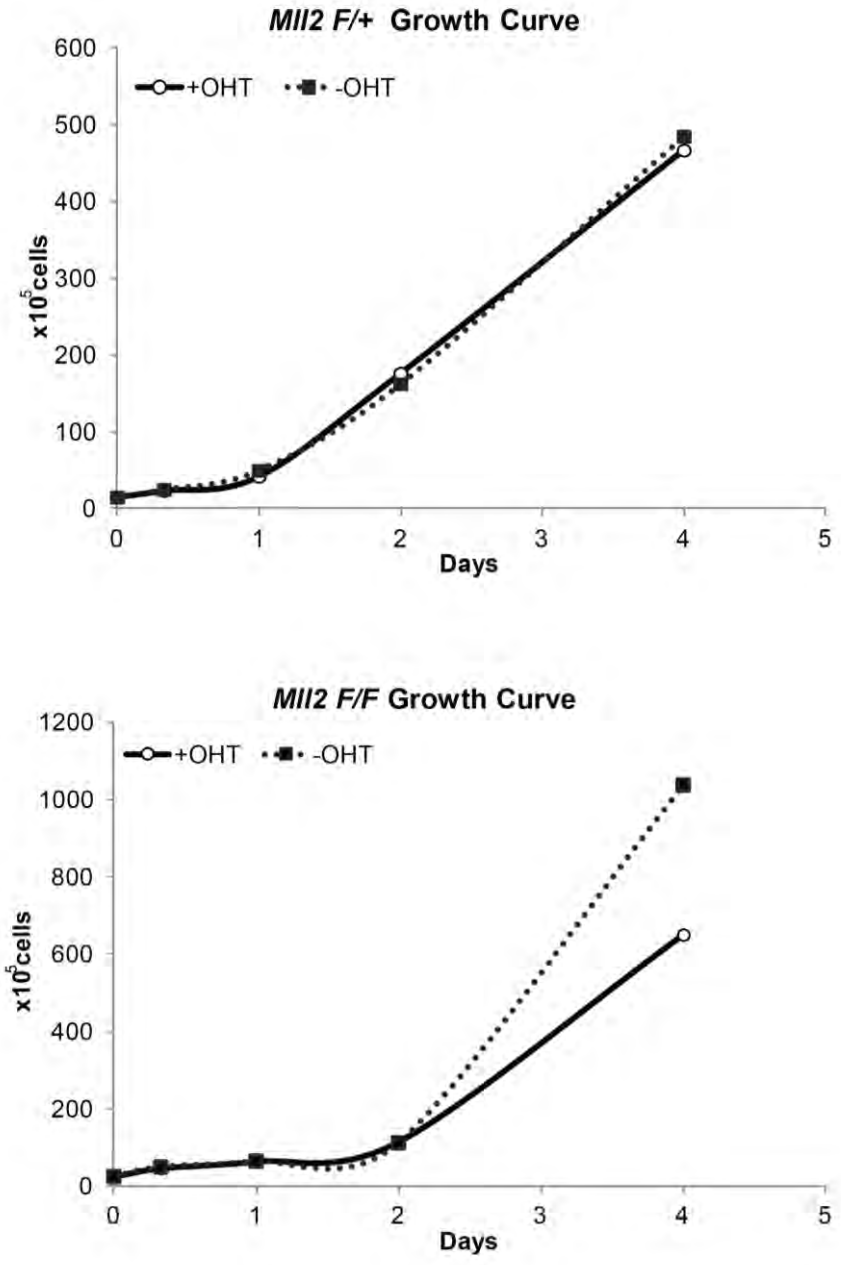

Figure 3.14. Deletion of both *Mll2* alleles results in growth inhibition. Cells were grown under standard conditions, harvested and counted in a haemocytometer at the indicated time-points. The cells were passed normally during the course of this experiment (on days 1 and 3) and the dilution factor is calculated in the results presented. A representative of 3 experiments is shown.

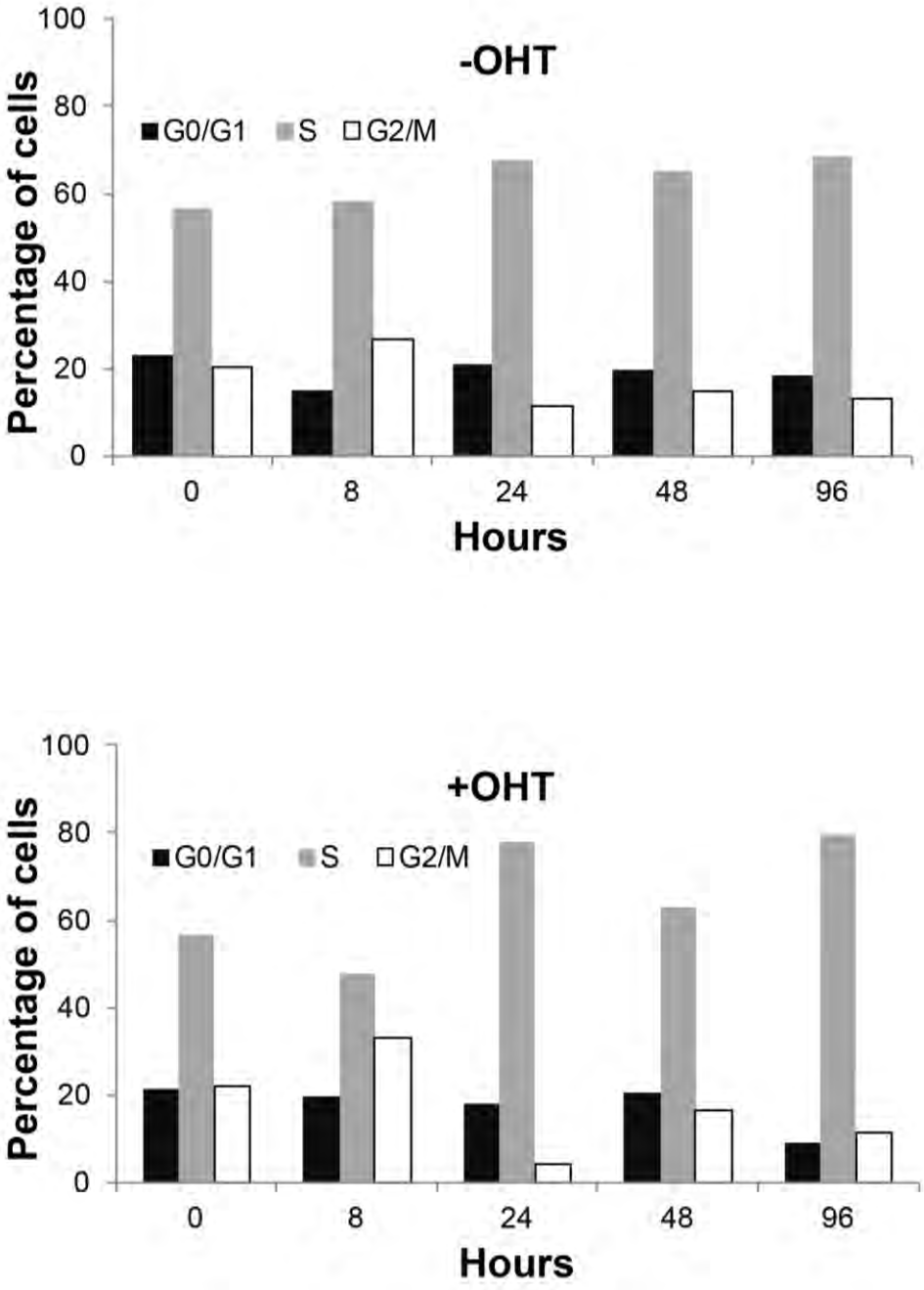

Figure 3.15. No apparent block in cell cycle progression after *Mll2* deletion. Cell cycle distribution of control (-OHT) and induced (+OHT) *Mll2*$^{F/F}$ cells was measured by propidium iodide staining and fluorescence activated cell sorting after methanol fixation over a 4 day time-course. A representative of 2 experiments is shown.

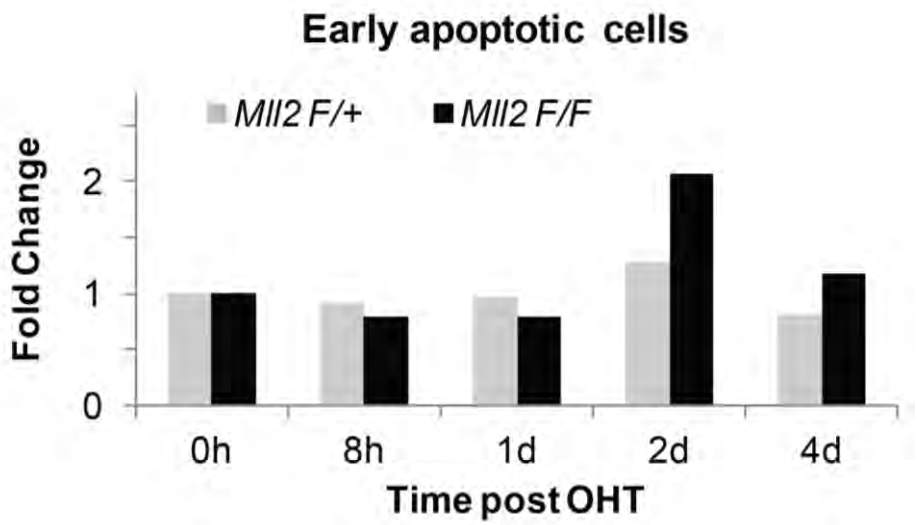

Figure 3.16. Increased apoptosis after MLL2 depletion. *Mll2*$^{F/F}$ and *Mll2*$^{F/+}$ cells were grown under standard conditions and induced with OHT. The cells were harvested at the indicated timepoints, stained with FITC conjugated Annexin V and propidium iodide (PI) and analysed by fluorescence activated cell sorting. Bars represent the mean fold change of the Annexin V positive, PI negative cell fraction (2 biological replicates).

3.2.3. *Magoh2* steady state mRNA levels decrease 2 days post OHT induction

To establish the kinetics of *Magoh2* transcriptional silencing we measured *Magoh2* steady state mRNA levels over a 4-day period post OHT induction (*Figure 3.17*). Although *Mll2* deletion was complete within 1 day post addition of OHT, the first transcriptional effects on *Magoh2* were only observed 2 days post induction. This discrepancy can be attributed to both the stability of the *Magoh2* mRNA and the half-life of the MLL2 protein, which has been shown to be approximately 36 hours in the same cell system (Glaser et al., 2009). Due to the lack of a suitable antibody the kinetics of MLL2 depletion could not be examined in this study and indirect effects cannot be completely excluded. It is however reasonable to assume that MLL2 levels reach critically low levels between 1 and 2 days post OHT induction, as this would be consistent with previously published reports (Glaser et al., 2009) and both an increase in apoptosis and a depletion of *Magoh2* mRNA were first observed at this time point. Subsequently, a nearly complete depletion of *Magoh2* mRNA was observed 4 days post OHT induction. No significant down-regulation of *Magoh2* was observed in the $Mll2^{F/+}$ cells, suggesting that one allele of *Mll2* was sufficient to maintain *Magoh2* expression. There was however no indication on whether MLL2 protein levels in the $Mll2^{F/+}$ cells were decreased by deletion of one *Mll2* allele or if the cells compensated for this

loss by increasing transcription or translation rates of the remaining *Mll2* allele.

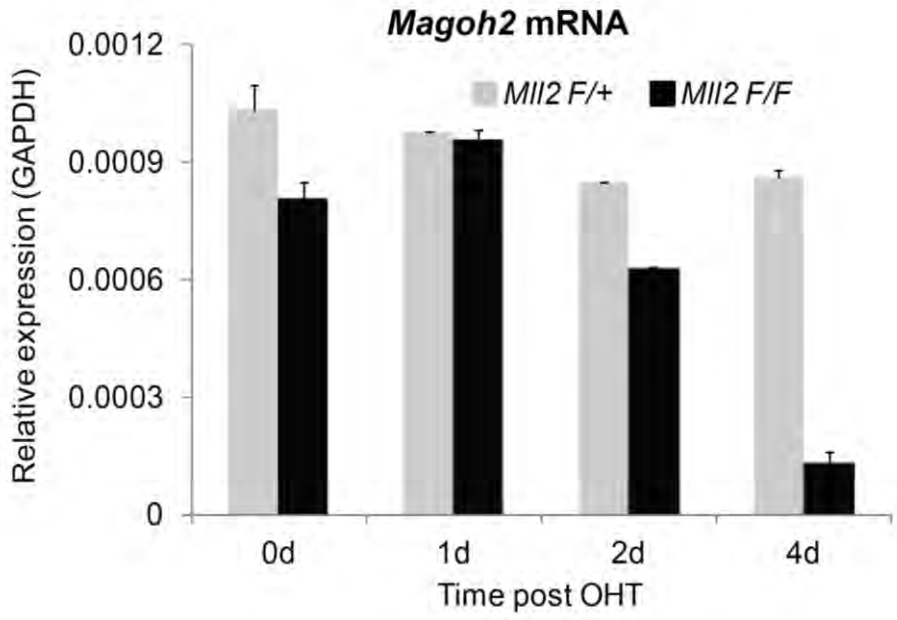

Figure 3.17. *Magoh2* is transcriptionally silenced 4 days post *Mll2* deletion. *Magoh2* mRNA expression after deletion of both alleles of *Mll2*, as detected by reverse transcription and quantitative PCR. Bars represent mean +sd of 4 measurements (2 biological x 2 technical)

3.2.4. The pre-initiation complex on the *Magoh2* promoter is disrupted 4 days post OHT induction

We have previously shown using *Mll2*$^{-/-}$ cells that the *Magoh2* promoter is not occupied by what we assume is a RNA polymerase II pre-initiation complex. To gain insight into the kinetics of RNA polymerase II depletion, we repeated the DNaseI footprinting experiments described previously, this time using the inducible knock-out system over a time-course of 8 days (*Figure 3.18*). We observed the characteristic DNaseI protection around the transcription start site flanked by hypersensitive regions, as described previously in E14 ES cells (*Figure 3.*10). This pattern was maintained in the *Mll2*$^{F/+}$ cells throughout this time-course, consistent with the reported maintenance of *Magoh2* mRNA levels. Contrastingly, the DNaseI digestion pattern observed in *Mll2*$^{F/F}$ cells over the *Magoh2* transcription start site was altered 4 days post OHT induction, coinciding with the observed *Magoh2* mRNA depletion. The digestion pattern observed 8 days post OHT induction very closely resembled the pattern seen previously in the *Mll2*$^{-/-}$ cells. These experiments point to a requirement for MLL2 in order to maintain stable association of the RNA polymerase II complex with the *Magoh2* promoter *in vivo*.

As DNaseI footprinting only detects DNaseI mediated cuts on one strand of the DNA molecule, the assay was repeated with primers specific for the opposite strand, verifying the results described above (*Figure 3.19*).

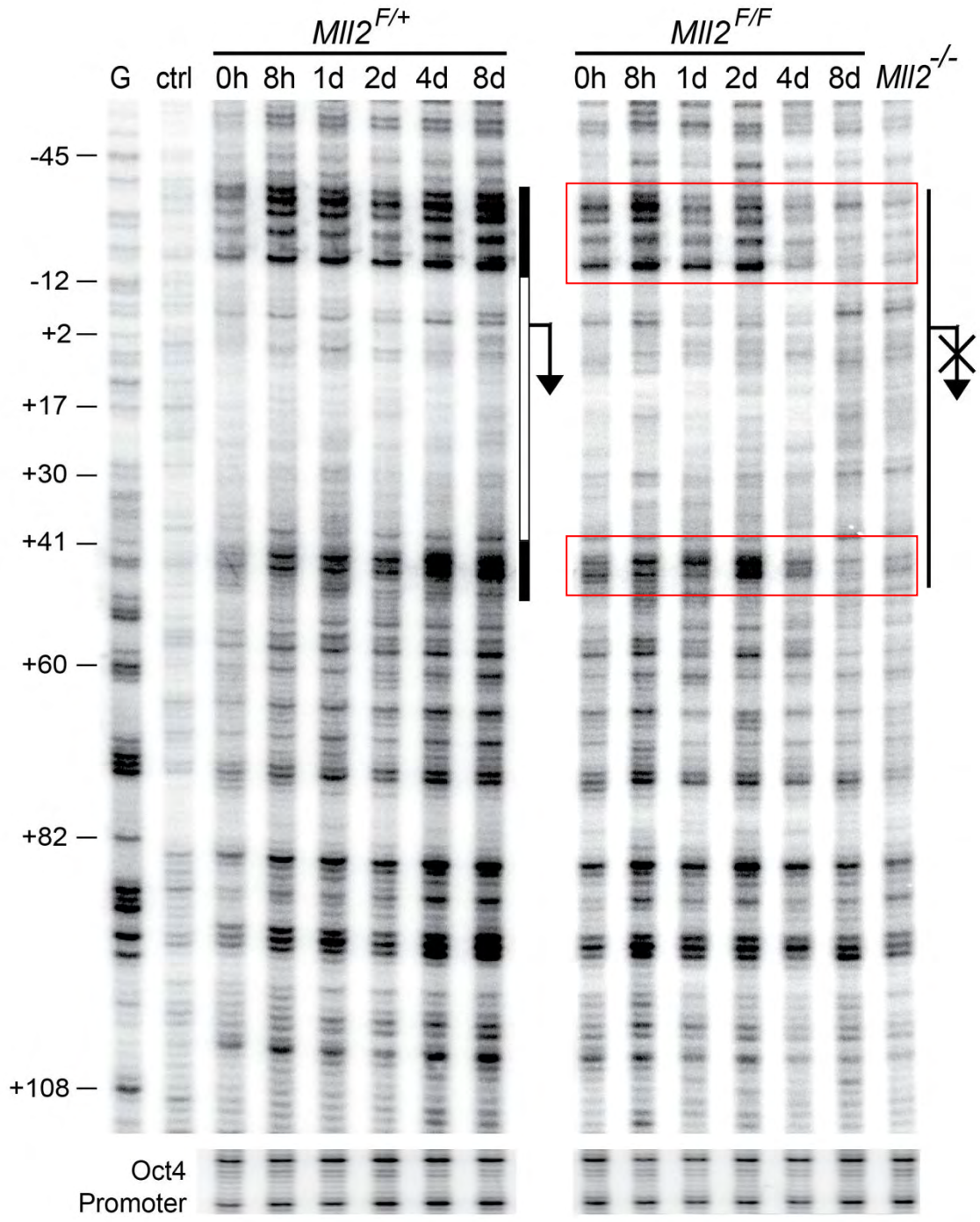

Figure 3.18. MLL2 is required to maintain a stable RNA polymerase II complex on the *Magoh2* promoter. DNaseI footprinting over the *Magoh2* promoter in OHT treated *Mll2*$^{F/F}$ and *Mll2*$^{F/+}$ cells. The region protected from DNaseI digestion spans from -12bp to +35bp, as indicated by the white bar. As the control sample appears under-loaded, protection is detected as decreased intensity as compared to the lane average signal intensity. The protected region is flanked by hypersensitive regions (black bars) G: G-reaction, ctrl: *in vitro* DNaseI digested DNA. Time post OHT induction is indicated above the relevant lanes. *Mll2*$^{-/-}$ ES cells were included to demonstrate the DNaseI digestion pattern in the inactive state. Arrow: transcription start site. Annotation is relative to the transcription start site. Primers on the *Oct4* promoter were used to verify equal DNaseI digestion of all samples. The regions that change in DNaseI sensitivity are shown in the red boxes.

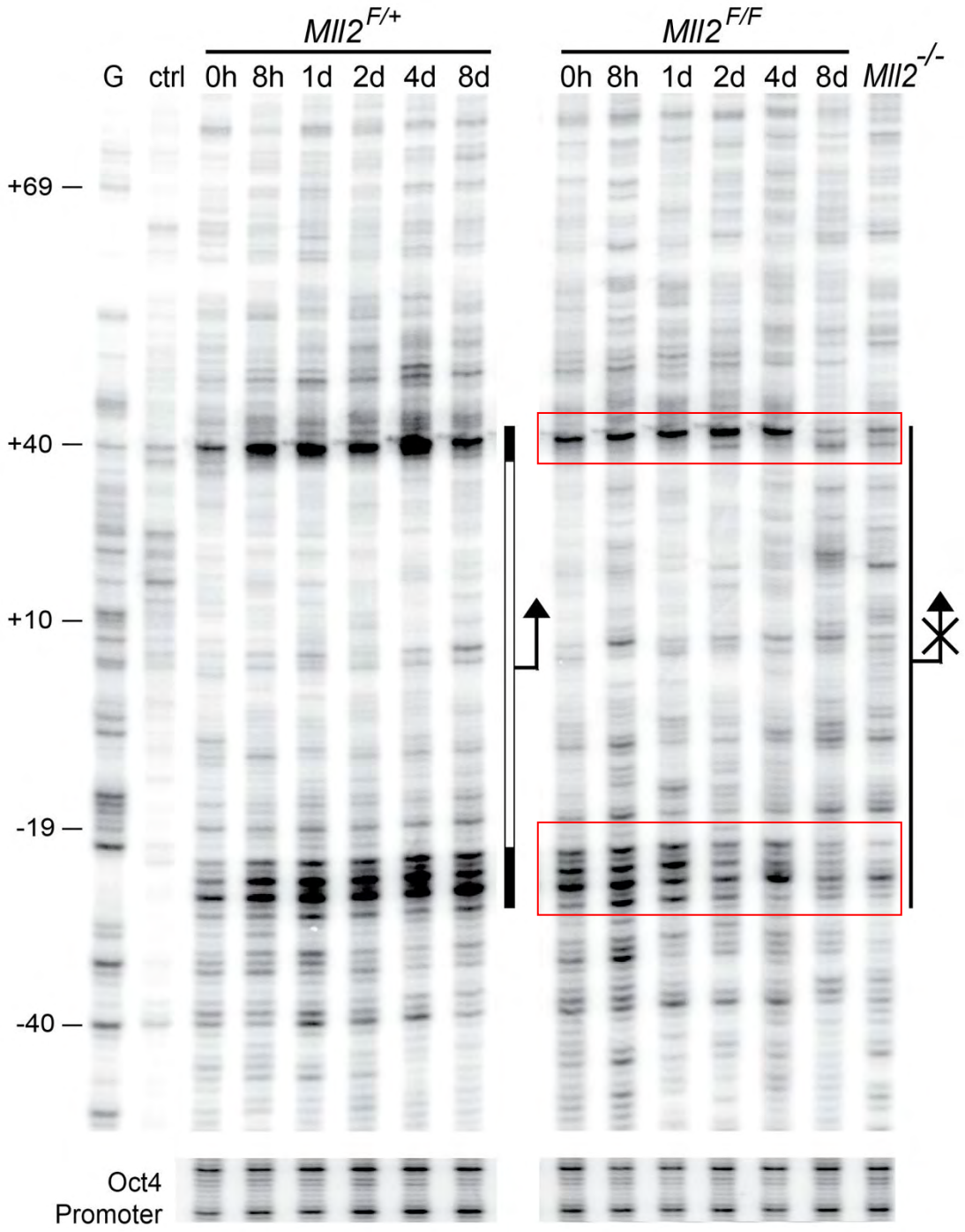

Figure 3.19. MLL2 is required to maintain a stable RNA polymerase II complex on the *Magoh2* promoter – analysis of reverse strand. A different set of primers was used on the reverse strand to verify the results illustrated in *Figure 3.18*. As the control sample appears under-loaded, protection is detected as decreased intensity as compared to the lane average signal intensity. Black bars indicate hypersensitive regions while white bars indicate protected regions. G: G-reaction, ctrl: *in vitro* DNaseI digested DNA. Time post OHT induction is indicated above the relevant lanes. *Mll2*$^{-/-}$ ES cells were included to demonstrate the DNaseI digestion pattern in the inactive state. Arrow: transcription start site. Annotation is relative to the transcription start site. Primers on the *Oct4* promoter were used to verify equal DNaseI digestion of all samples. The regions that change in DNaseI sensitivity are shown in the red boxes.

3.2.5. Loss of the active histone marks and RNA polymerase II closely follow *Magoh2* expression levels

To further examine the kinetics of chromatin alterations accompanying *Magoh2* silencing we employed chromatin immunoprecipitation over a 4-day time-course after *Mll2* deletion (*Figure 3.20*). In concordance with the results presented above, loss of the active H3K4me$_3$ and H3K9ac marks was first observed 2 days post OHT induction in *Mll2*$^{F/F}$ cells. The levels of both histone marks decreased to near background levels 4 days post induction, concomitantly with *Magoh2* mRNA depletion. RNA polymerase II levels closely correlated with the levels of these two histone marks. Loss of only one *Mll2* allele (*Mll2*$^{F/+}$ cells) had no effect on the chromatin environment of the *Magoh2* promoter. These experiments are consistent with the view that MLL2 and/or the H3K4me$_3$ mark are a prerequisite for RNA polymerase II recruitment to the *Magoh2* promoter.

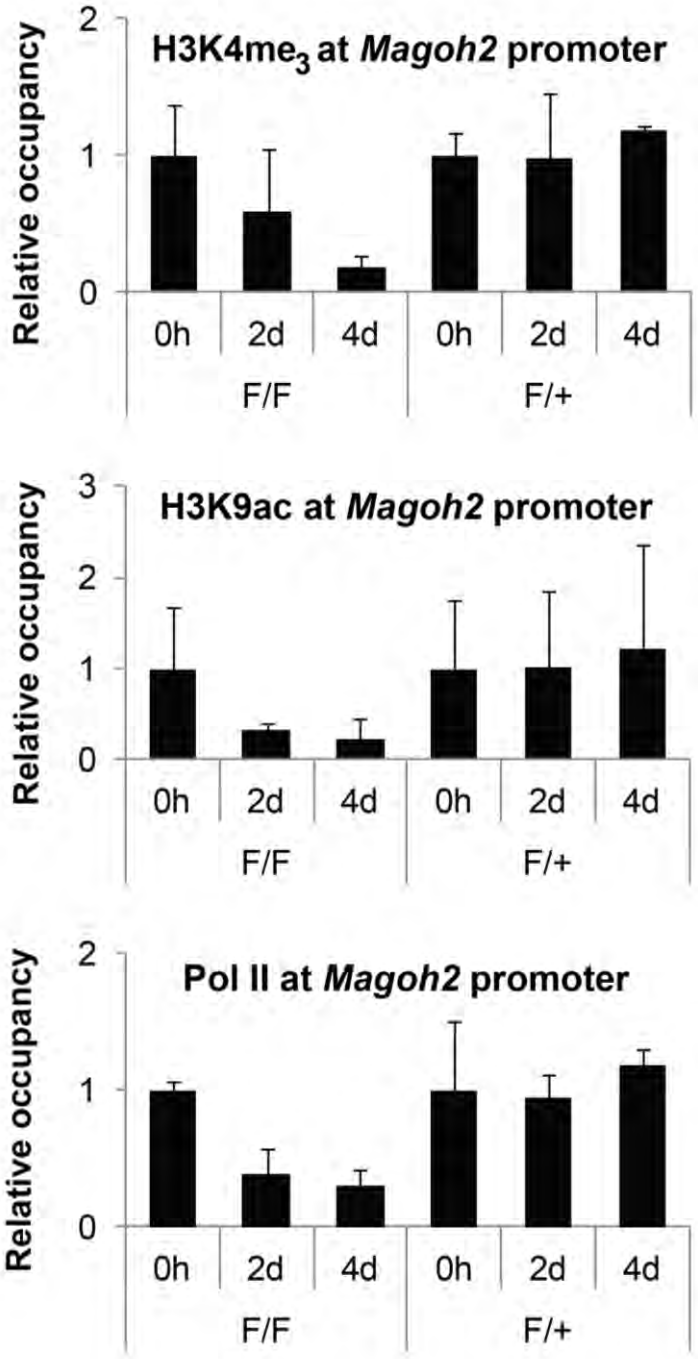

Figure 3.20. Loss of active chromatin marks and RNA polymerase II on the *Magoh2* promoter up to 4 days post *Mll2* deletion. Chromatin immunoprecipitation assays were performed using antibodies against H3K4me$_3$ **(top)**, H3K9ac **(middle)** and the RNA polymerase II CTD **(lower)**. The immunoprecipitated DNA was amplified by quantitative PCR with a primer set recognizing the *Magoh2* promoter (the amplicon is centred around +60bp). Data for histone modifications are corrected for nucleosome density as determined by a histone H3 immunoprecipitation. Data for RNA polymerase II are corrected for the signal obtained with input material. All the data presented are normalized against the signal obtained with primers for the *Oct4* promoter and plotted as fold change over the 0 hour time-point (untreated cells)

3.2.6. Nucleosome remodelling takes place on the *Magoh2* promoter as soon as RNA polymerase II is removed

DNaseI hypersensitive sites and active promoters have been reported to be relatively depleted of nucleosomes. In order to investigate whether *Magoh2* silencing is accompanied by changes in promoter nucleosome occupancy we employed Micrococcal nuclease (MNase) footprinting after deletion of *Mll2* in $Mll2^{F/F}$ and $Mll2^{F/+}$ cells. MNase primarily catalyses double strand DNA breaks. Its action is greatly inhibited by the presence of nucleosomes or other DNA bound proteins, thus protection from MNase digestion after *in vivo* MNase treatment reflects regions that are occupied by nucleosomes or other protein complexes in live cells, while MNase sensitive regions are naked DNA (linker DNA or nucleosome free regions) (Axel, 1975; Clark and Felsenfeld, 1971; Zhang and Gralla, 1989). Identical MNase digestion patterns were observed in untreated $Mll2^{F/F}$ and $Mll2^{F/+}$ cells. No changes were observed in $Mll2^{F/+}$ cells up to 8 days post OHT induction. Contrastingly, in $Mll2^{F/F}$ cells the first changes were evident in the immediate region of the *Magoh2* transcription start site 2 days post OHT induction (*Figure 3.21*). We observed a decrease in MNase cutting frequency in the previously hypersensitive regions and an increase in MNase cutting frequency in regions that were previously almost completely protected. Changes in MNase accessibility in $Mll2^{F/F}$ cells were also observed downstream of the transcription site 2 days post OHT induction (*Figure 3.22*). These results demonstrate that 2 days post OHT induction,

nucleosomes and/or other proteins are being repositioned on the *Magoh2* promoter.

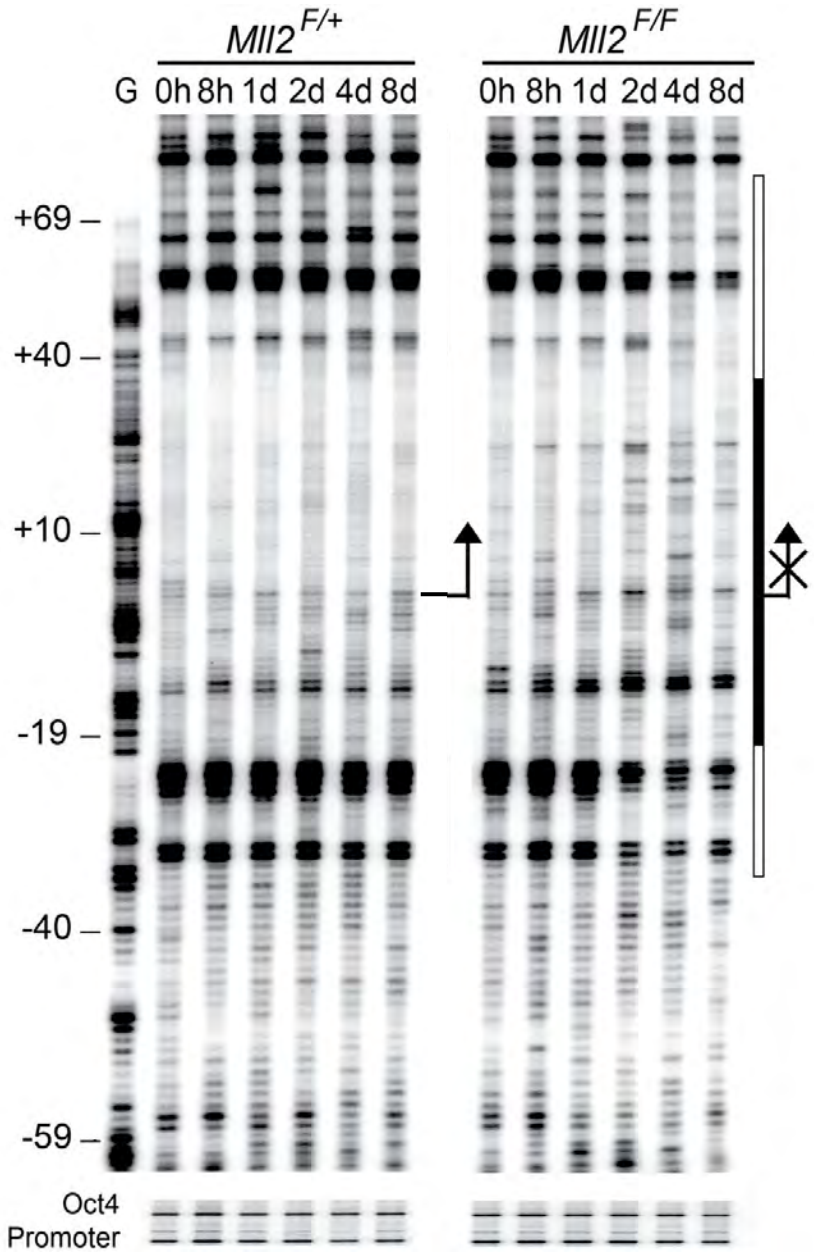

Figure 3.21. Nucleosome remodelling over the *Magoh2* promoter up to 8 days post *Mll2* deletion. Micrococcal nuclease (MNase) footprinting in the vicinity of the *Magoh2* transcription start site in OHT treated *Mll2*$^{F/F}$ cells. Black bars indicate hypersensitive regions while white bars indicate protected regions when comparing digestion patterns observed at the 8 days time-point to the 0 hours time-point. G: G-reaction. Time post OHT induction is indicated at the top of each lane. Arrow: transcription start site. Annotation is relative to the transcription start site. Primers on the *Oct4* promoter were used to verify equal MNase digestion of all samples.

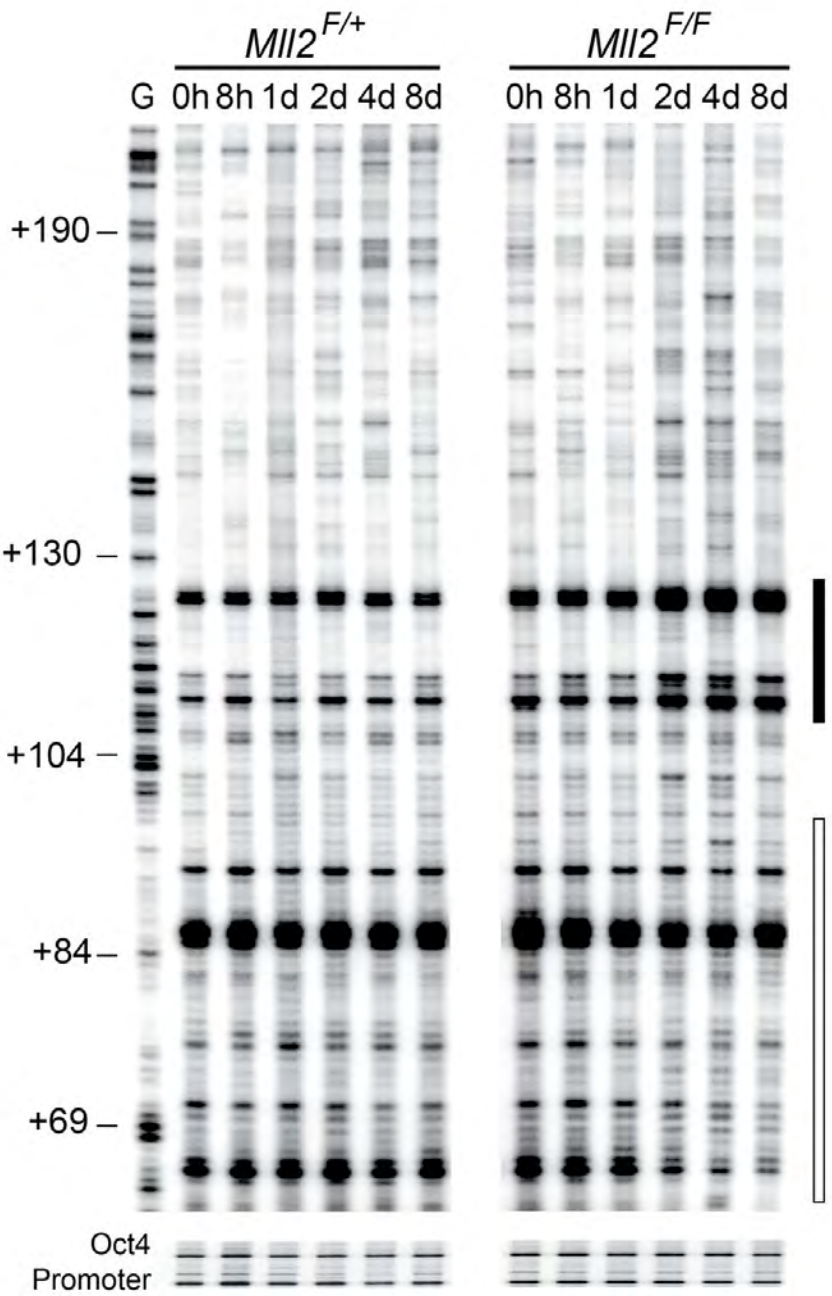

Figure 3.22. Nucleosome remodelling extends further downstream of the *Magoh2* transcription start site. MNase footprinting examining the *Magoh2* gene body. The black bar indicates a hypersensitive region while the white bar indicates a protected region when comparing digestion patterns observed at the 8 days time-point to the 0 hours time-point. G: G-reaction. Time post OHT induction is indicated at the top of each lane. Arrow: transcription start site. Annotation is relative to the transcription start site. Primers on the *Oct4* promoter were used to verify equal MNase digestion of all samples.

3.2.7. DNA methylation occurs shortly after transcriptional silencing

DNA methylation of CpG island promoters is one of the best characterised silencing marks. We have previously shown that the *Magoh2* promoter is methylated in the absence of MLL2. To investigate whether DNA methylation is a silencing initiating or maintaining event, we measured DNA methylation levels on the *Magoh2* promoter region by bisulphite conversion and pyrosequencing over an 8-day time-course in the *Mll2* inducible knock-out system (*Figure 3.23*). A total of 19 CpG sites distributed between -48bp and +69bp from the transcription start site were examined in this assay. The results demonstrated that DNA methylation increases in $Mll2^{F/F}$ cells 4 days post OHT induction, which was the time-point where the H3K4me$_3$ and H3K9ac histone marks along with RNA polymerase II had already been removed from the *Magoh2* promoter and *Magoh2* transcription had ceased. Thus, DNA methylation was not responsible for initiating *Magoh2* silencing in those cells. In $Mll2^{F/+}$ cells only a very small increase in DNA methylation was observed 8 days after OHT induction. As no clear bias towards methylation (or protection from methylation) of specific CpG dinucleotides was observed, the three-dimensional graphs presented in *Figure 3.23* were summarised in a simpler box-plot format (*Figure 3.24*). This representation reveals a significant (at 95% confidence) increase in CpG methylation on the *Magoh2* promoter in $Mll2^{F/F}$ cells 4 days post OHT treatment. At the 8 days

time-point CpG methylation on the *Magoh2* promoter in *Mll2*$^{F/F}$ cells exceeded the levels measured in *Mll2*$^{-/-}$ cells.

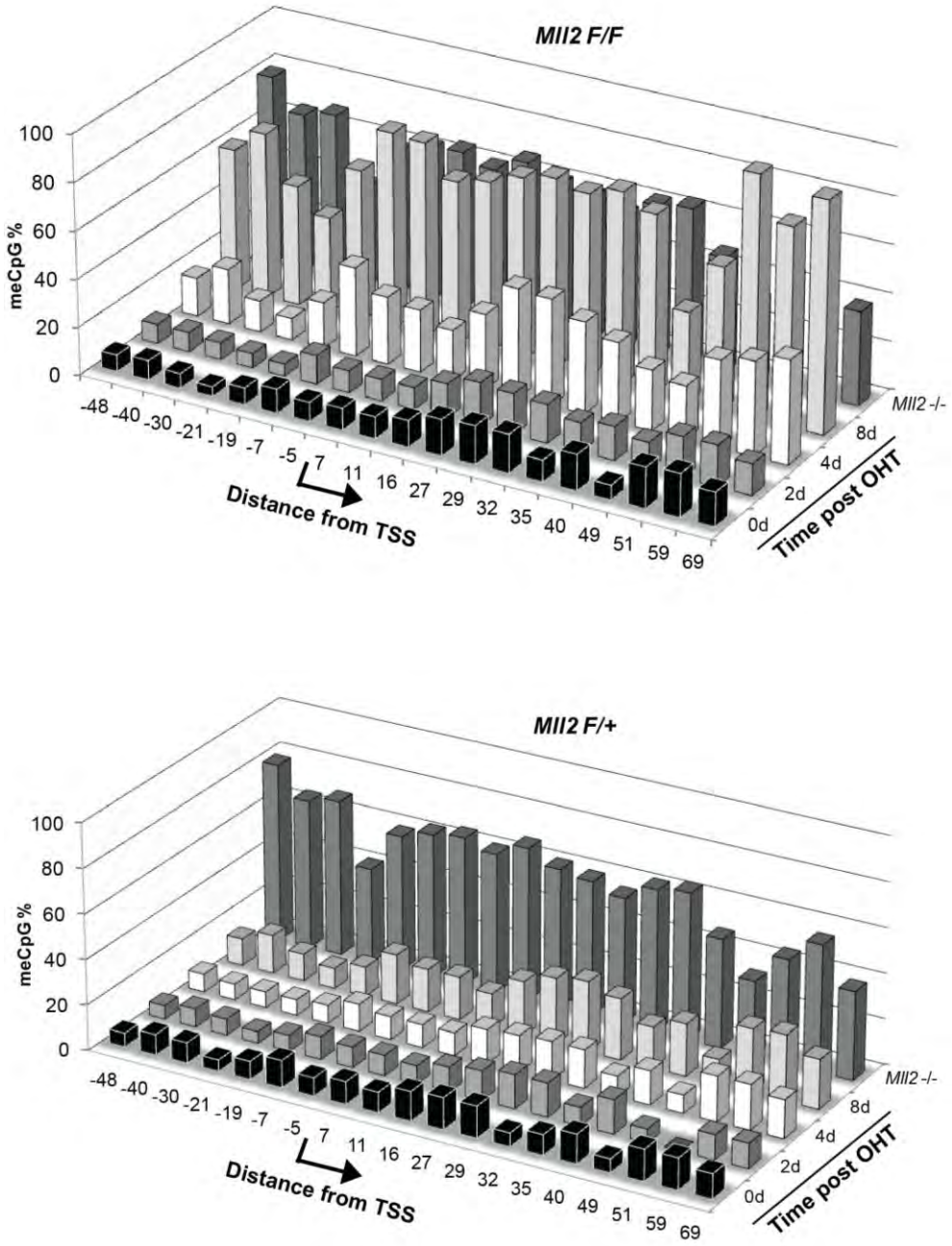

Figure 3.23. Extensive DNA methylation on the *Magoh2* CpG island in the absence of MLL2. DNA methylation levels on a part of the *Magoh2* CpG island promoter in *Mll2*$^{F/F}$ and *Mll2*$^{F/+}$ cells. The cells were induced with OHT and harvested at the time points indicated on the Z-axis. The DNA methylation levels measured in *Mll2*$^{-/-}$ cells (*Figure 3.12*) are inserted as a separate point on the Z-axis for comparison. Each point on the X-axis represents an individual CpG dinucleotide. **Top panel:** Methylation changes over a time-course after *Mll2* deletion in *Mll2*$^{F/F}$ cells **Lower panel:** DNA methylation measured at different time points in *Mll2*$^{F/+}$ cells up to 8 days post OHT induction.

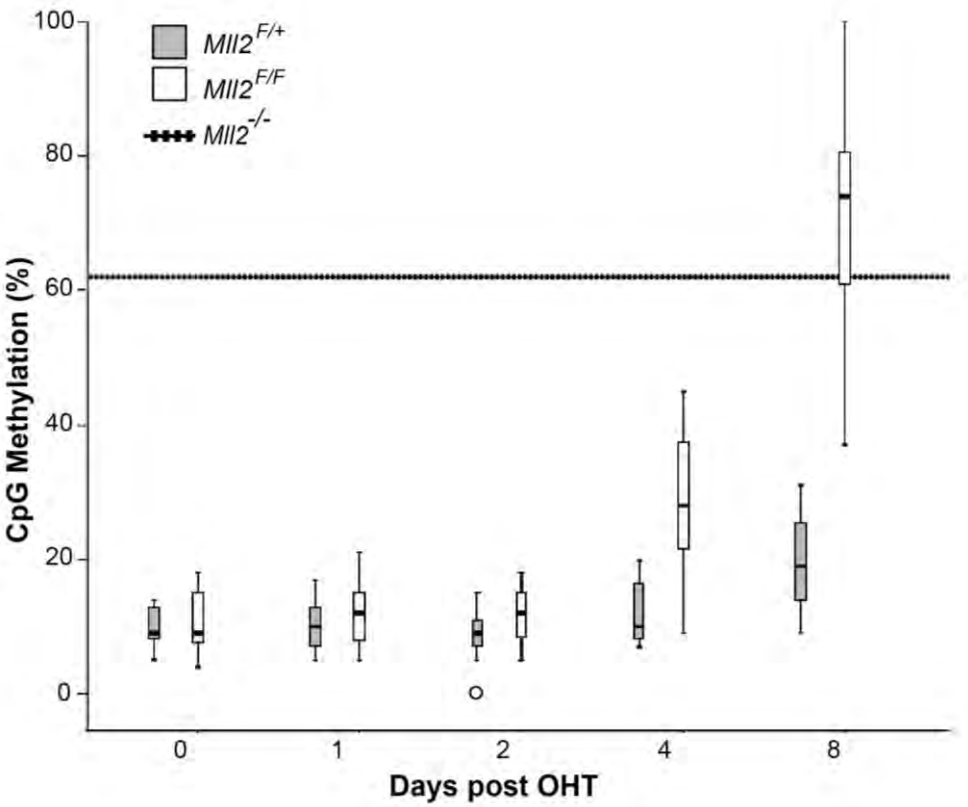

Figure 3.24. Summary of DNA methylation on the *Magoh2* CpG island promoter. The data depicted in *Figure 3.23* were summarized in a box-plot format for easier comparison. The hatched line represents the median DNA methylation level in the *Mll2*$^{-/-}$ cells. Black bars: median CpG methylation over the region examined (-48bp to +69bp). Boxes: 95% confidence interval, Whiskers 99% confidence interval. Open circle: outlier.

3.2.8. Active transcription and RNA polymerase II are not required for the maintenance of the H3K4me$_3$ mark or for protection from DNA methylation

To investigate whether active transcription or the presence of RNA polymerase II were required to prevent DNA methylation, *Mll2*$^{F/F}$ cells were treated with DRB or α-amanitin. DRB blocks transcriptional elongation by inhibiting p-TEFb enzymatic activity (Fraser et al., 1978; Marshall et al., 1996). α-amanitin binds the active site of RNA polymerase II and this complex is subsequently targeted for degradation (Nguyen et al., 1996). Whole cell extract Western blots demonstrated that serine 2 phosphorylated elongating form of RNA polymerase II was depleted after 16 hours of DRB treatment, while a 24 or 48 hour α-amanitin treatment greatly reduced total RNA polymerase II levels (*Figure 3.25 Top panel*. An antibody against nucleoporin Nup188 was used to verify similar loading of all samples.

To correlate these findings with *Magoh2* expression, we measured *Magoh2* transcript levels in DRB and α-amanintin treated cells by RT-QPCR. As expected, primary transcript and steady state mRNA levels were barely detectable following treatment with either of the two drugs (*Figure 3.25*). To verify this result, GAPDH expression levels were measured as well and were found to be just above detection limits (*data not shown*). Importantly, 18S rRNA could be readily detected as RNA polymerase I is reported to be unaffected by either of these drugs.

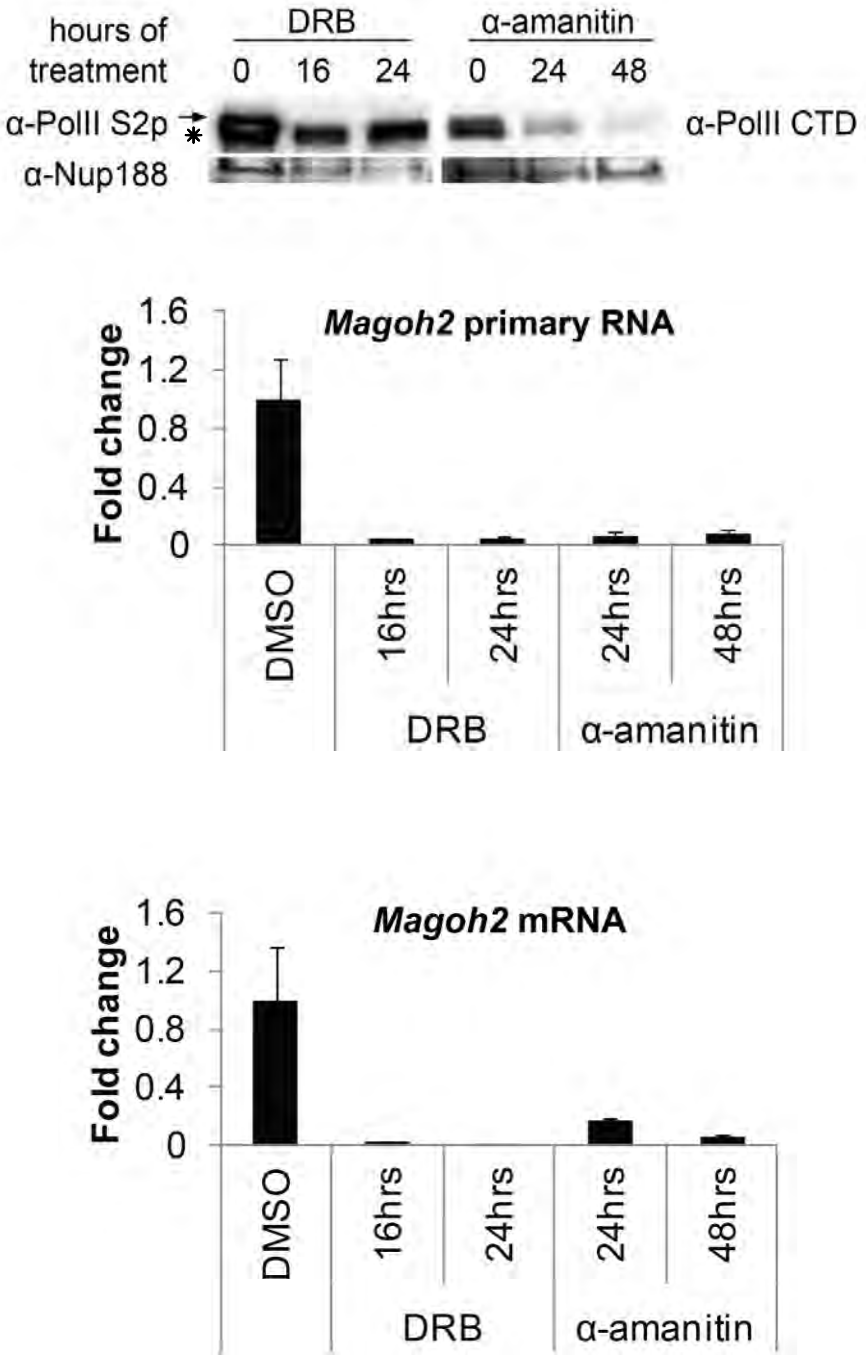

Figure 3.25. Treatment of *Mll2*$^{F/F}$ cells with DRB and α-amanitin completely blocked RNA polymerase II dependent transcription. Top panel: Western blot analysis on DRB and α-amanitin treated *Mll2*$^{F/F}$ cells, using whole cell protein extracts. The arrow indicates the band for phospho-serine 2 RNA polymerase II, the asterisk indicates a non-specific band (possibly non-modified Pol II). DMSO treated samples are used as a control to exclude drug carrier effects. **Middle panel:** RT-QPCR detecting primary *Magoh2* RNA in DRB/α-amanitin treated cells. Data presented here are normalized to the 18S rRNA which is transcribed by RNA polymerase I. Bars represent mean +sd of 4 measurements (2 biological x 2 technical replicates). **Lower panel:** RT-QPCR detecting *Magoh2* mRNA in DRB/α-amanitin treated cells. Data presented here are normalized to the 18S rRNA. Bars represent mean +sd of 4 measurements (2 biological x 2 technical replicates).

Having established that *Magoh2* transcription had ceased, we examined chromatin state at the *Magoh2* promoter by chromatin immunoprecipitation (*Figure 3.26*). No change in RNA polymerase II occupancy or H3K4me$_3$ levels was observed in DRB treated cells. On the contrary, cells treated with α-amanitin exhibited approximately 50% reduction in RNA polymerase II occupancy, accompanied by approximately 2-fold increase in H3K4me$_3$ levels over the *Magoh2* promoter. These findings are consistent with the notion that H3K4me$_3$ deposition and maintenance on the *Magoh2* promoter does not depend on RNA polymerase II recruitment and transcription.

To identify what maintains the *Magoh2* CpG island in an un-methylated state, we examined the DNA methylation levels on the *Magoh2* promoter after DRB induced transcriptional blocking or α-amanitin induced RNA polymerase II depletion by HpaII digestion and QPCR (*Figure 3.26*). We did not observe any increase in methylation of this particular CpG dinucleotide following either treatment. This experiment suggests that in the presence of MLL2, protection from DNA methylation does not depend on productive transcriptional elongation and is not affected by a severe depletion of RNA polymerase II binding. The presence of MLL2 and/or the H3K4me$_3$ mark appear to be sufficient to maintain the *Magoh2* promoter free from methylation.

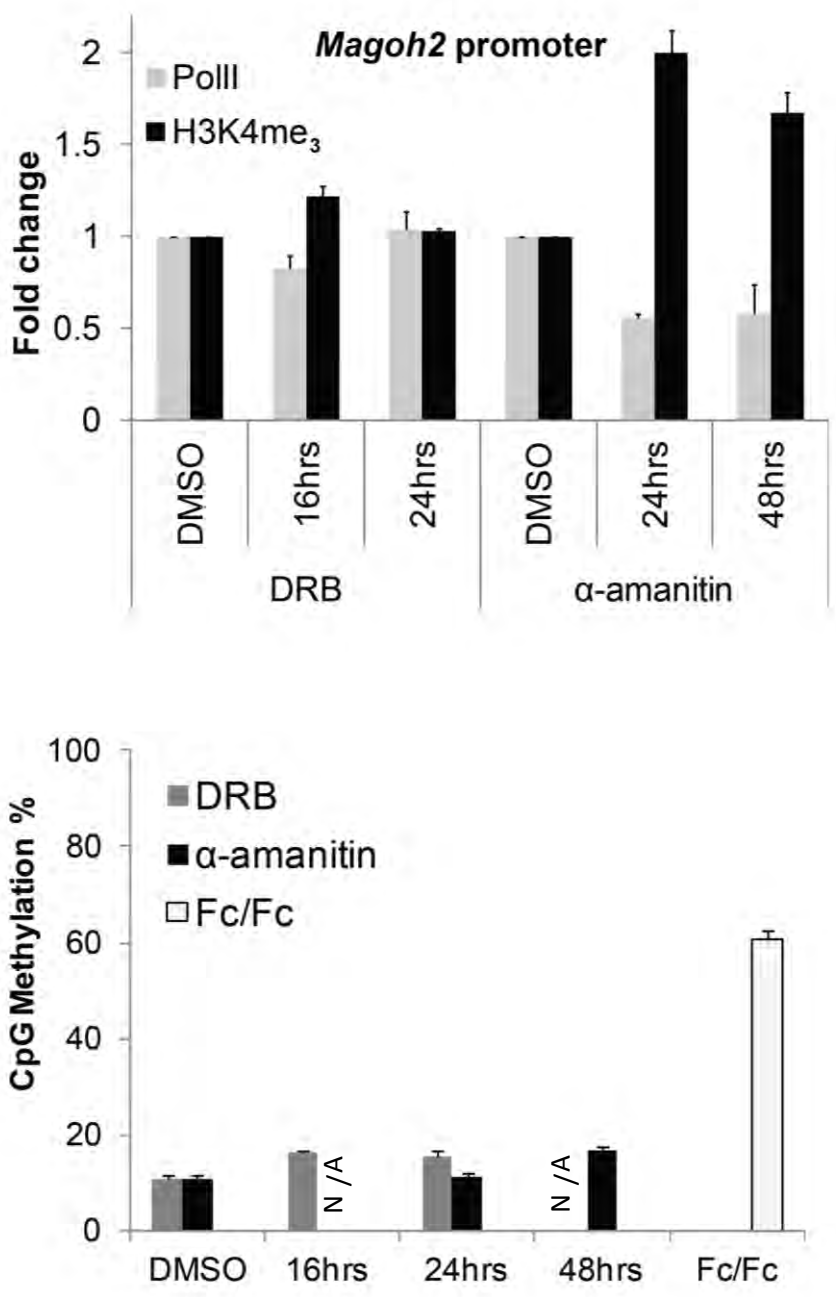

Figure 3.26. Depletion of RNA polymerase II does not affect DNA methylation levels and results in an increase in H3K4me$_3$ levels on the *Magoh2* promoter. Top panel: ChIP with antibodies against H3K4me$_3$ and RNA polymerase II CTD. Data for histone modifications are corrected for nucleosome density as determined by a histone H3 immunoprecipitation. Data for RNA polymerase II are corrected for input. The data presented are plotted as fold change over the DMSO mock-treated cells. Bars represent mean +sd of 4 measurements (2 biological x 2 technical replicates). **Lower panel:** DNA methylation levels were measured by QPCR with primers spanning a single HpaII site after HpaII digestion. Methylation levels of the same CpG dinucleotide 8 days post OHT treatment of *Mll2*$^{F/F}$ cells (Fc/Fc) are shown for comparison. Data presented here are normalized over an amplicon on the *Oct4* promoter that does not contain a HpaII site. N/A: not assayed. Bars represent mean +sd of 4 measurements (2 biological x 2 technical replicates). DMSO treated cells are used to exclude any drug carrier effects.

3.3. Reactivation of the *Magoh2* promoter

3.3.1. Re-expression of *Mll2* from the endogenous locus

In order to examine whether *Magoh2* silencing was reversible or DNA methylation prevented reactivation of the silenced CpG island promoter, we reactivated endogenous *Mll2* alleles in *Mll2*$^{-/-}$ ES cells. As described previously, the *Mll2* gene in these cells has been modified by the addition of a gene-trapping stop cassette in intron 1 (*Figure 3.27*). The *Mll2*$^{-/-}$ ES cells were transfected with a FlpO expressing plasmid (pCAGGS-FlpO-IRES-puro, courtesy of Dr. F. Stewart) and co-transfected with a 100-fold less GFP expressing plasmid (pMAX-GFP, Amaxa). The transfected cells were enriched by culturing in 0.5µg/ml or 1µg/ml puromycin containing medium. FACS analysis demonstrated that over 95% of the surviving cells expressed GFP (*data not shown*). Expression of the FlpO recombinase resulted in recombination between the two FRT sites flanking the stop cassette and removal of the cassette. Effectively this recombination event converts the *Mll2*$^{-}$ allele into *Mll2*^{F}.

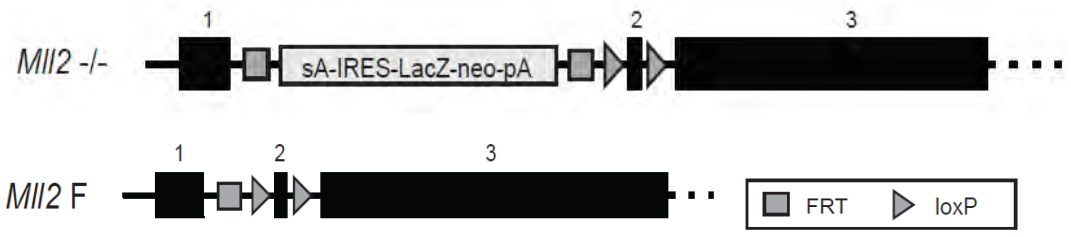

Figure 3.27. Reactivation of the endogenous *Mll2* alleles in *Mll2*$^{-/-}$ ES cells. The stop cassette is excised by FlpO mediated recombination allowing transcription of full length *Mll2*. Black boxes indicate Mll2 exons.

Successful deletion of the stop cassette and reactivation of *Mll2* were verified by genotyping PCR (*data not* shown) and *Mll2* steady state mRNA expression measurement (*Figure 3.28*) respectively. When *Mll2* was re-expressed, *Magoh2* mRNA could be detected at approximately wild type levels, suggesting that the epigenetic silencing described above can be counteracted by re-introduction of MLL2. Importantly, cells transfected with the GFP expressing plasmid alone (mock transfected) exhibited no *Mll2* or *Magoh2* expression.

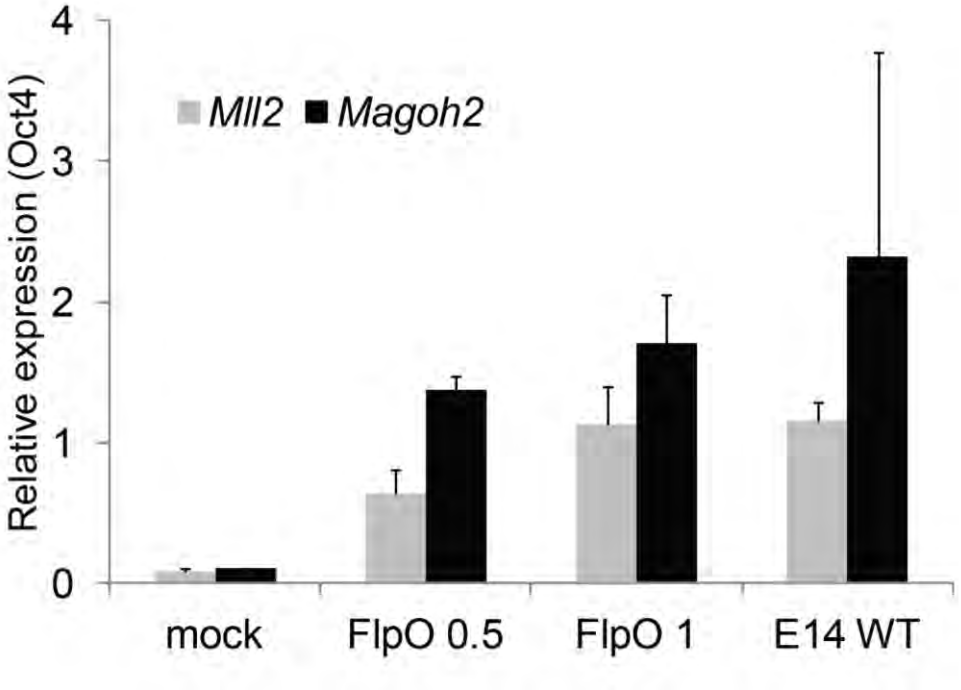

Figure 3.28. Reactivation of the endogenous *Mll2* rescues *Magoh2* expression. The FlpO transfected *Mll2*$^{-/-}$ cells were cultured in the presence of 0.5µg/ml or 1µg/ml puromycin. Both concentrations were sufficient to kill mock transfected cells, however only the cells overexpressing FlpO and therefore the puromycin resistance gene survive the higher concentration of antibiotic. Bars represent mean +sd of at least 4 measurements (2 biological x 2 technical replicates).

3.3.2. *Magoh2* expression is MLL2 dose-dependent

We previously showed that re-expression of the endogenous MLL2 is sufficient to at least partially reactivate the silenced *Magoh2*. We now wanted to test whether increased *Mll2* expression levels would lead to a more robust expression of *Magoh2.* To this end, $Mll2^{F/F}$ and $Mll2^{Fc/Fc}$ were transfected with a BAC expressing a MLL2-GFP fusion protein and a neomycin resistance gene (provided by Dr. H. Hofemeister, Stewart lab, *Figure 3.29 Top panel*). We established stable transgenic lines by culturing the transfected cells in the presence of G418 for 21 days in total. Within that time, emerging resistant colonies were isolated and cultured separately to yield individual clones. The BAC transfected $Mll2^{F/F}$ cells were subsequently treated with OHT to delete the endogenous *Mll2* alleles. We verified that the BAC transgenes express functional MLL2, as detected by *Mll2* mRNA expression and maintenance of *Magoh2* mRNA expression (*Figure 3.29 Middle panel*). The BAC transfected $Mll2^{F/F}$ cells will be hereafter referred to as “loss of genomic *Mll2*” (LoG) cells while the BAC transfected $Mll2^{Fc/Fc}$ cells as “Rescue” cells.

Mll2 and *Magoh2* mRNA expression were measured in two LoG clones and five Rescue clones (*Figure 3.29 Middle panel*). *Magoh2* expression was maintained in both LoG clones despite the deletion of the endogenous *Mll2* alleles, demonstrating that the BAC transgenes encoded functional MLL2. *Magoh2* expression levels closely followed those of *Mll2* pointing to an MLL2 dose-dependent effect on *Magoh2* transcription.

Rescue clones 1-3 showed only minimal expression of *Mll2* and no reactivation of *Magoh2*. Rescue clone 4 only exhibited partial *Magoh2* reactivation despite expressing wild type levels of *Mll2*. This can be explained if most of the BAC integrations only contain the 3'-most part of the BAC that may be expressed from a cryptic promoter. This would result in detection of this truncated RNA transcript but not in production of full length MLL2. Rescue clone 5 supports the reversibility of *Magoh2* silencing – which we observed before by re-activation of endogenous Mll2 (*Figure 3.28*) – and the MLL2 dose-dependent effect on *Magoh2* expression as the approximately 3-fold overexpression of *Mll2* is accompanied by a similar overexpression of *Magoh2.*These findings further support the idea that MLL2 acts upstream of transcriptional initiation and its increased presence facilitates the assembly of a functional RNA polymerase II complex on the *Magoh2* promoter.

To examine whether reactivation is mediated by re-establishment of H3K4me$_3$ we performed ChIP with anibodies against this modification (*Figure 3.29 Lower* panel). Data presented for the LoG cells are the mean enrichment obtained for clones 1 and 2 over a time period of 8 days post OHT induction. BAC derived MLL2 was sufficient to maintain H3K4 methylation on the *Magoh2* promoter after the endogenous *Mll2* was deleted. The H3K4me$_3$ levels however did not strictly correlate with *Magoh2* expression levels. Only a small difference in H3K4me$_3$ levels was observed between LoG clones 1 and 2 (as illustrated by the size of the error bars), despite the over 4-fold difference in *Magoh2* expression. This

finding suggests that MLL2 may have an H3K4 methylation-independent role in transcriptional regulation. Consistent with the low *Magoh2* expression levels measured in rescue clone 4, H3K4me$_3$ levels were found to be lower than normal supporting the view that the *Mll2* RNA measured previously in this cell line does not code for functional MLL2. Rescue clone 5 exhibited a complete rescue of the H3K4me$_3$ mark on the *Magoh2* promoter upon re-introduction of MLL2.

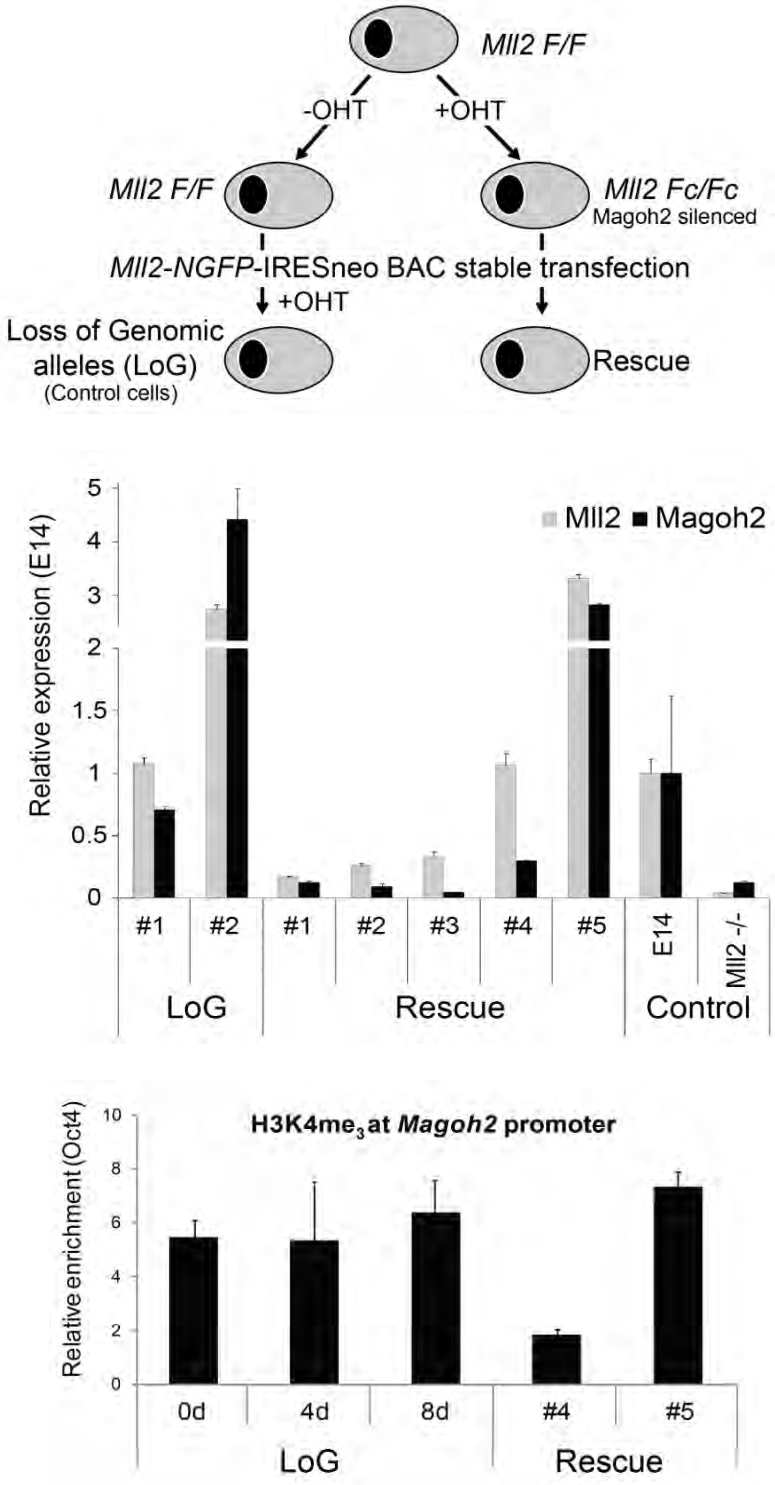

Figure 3.29. MLL2 has a dose-dependent effect on *Magoh2* transcription. Top panel: *Mll2*$^{F/F}$ and *Mll2*$^{Fc/Fc}$ cells were transfected with an *Mll2* expressing BAC (*Mll2*-NGFP-IRESneo). **Middle panel:** Expression analyses in LoG +OHT, Rescue, E14 and *Mll2*$^{-/-}$ cells. Bars represent the mean +sd of 4 measurements (2 biological x 2 technical replicates). Expression levels are normalized against GAPDH expression and illustrated as expression fold-change over the expression levels in E14 wild type cells. **Lower panel:** H3K4me$_3$ chromatin immunoprecipitation on the *Magoh2* promoter. Data presented for the LoG cells show the mean +sd H3K4me$_3$ levels in LoG clones 1 and 2. Bars for the Rescue clones represent the mean +sd of 4 measurements (2 biological x 2 technical replicates)

3.3.3. The entire MLL2 protein is required for maintenance of *Magoh2* expression

Although we showed that over-expression of *Mll2* can lead to a similar over-expression of *Magoh2,* it was not clear whether this was due to SET-domain mediated enzymatic activity. To try and identify whether the SET domain and/or other domains of MLL2 are required for *Magoh2* expression we measured *Magoh2* mRNA levels in $Mll2^{F/F}$ cells transfected with mutant-MLL2 BACs (cell lines generated and provided by Mr. A. Gupta and Mr. D.C. Torres, Stewart lab), before and after OHT induction. *Figure 3.30 Top* illustrates the MLL2 domains that were deleted in each of these cell lines. The results demonstrated that none of these mutant forms of MLL2 was able to maintain wild type *Magoh2* expression levels after the endogenous *Mll2* alleles had been deleted (*Figure 3.30 Lower panel*). Surprisingly, all the mutant-MLL2 expressing cell lines exhibited slightly reduced *Magoh2* expression even in the presence of wild type MLL2. Most notably, *Magoh2* mRNA levels in the ΔSNL mutant were close to background levels, as measured in $Mll2^{-/-}$ cells.

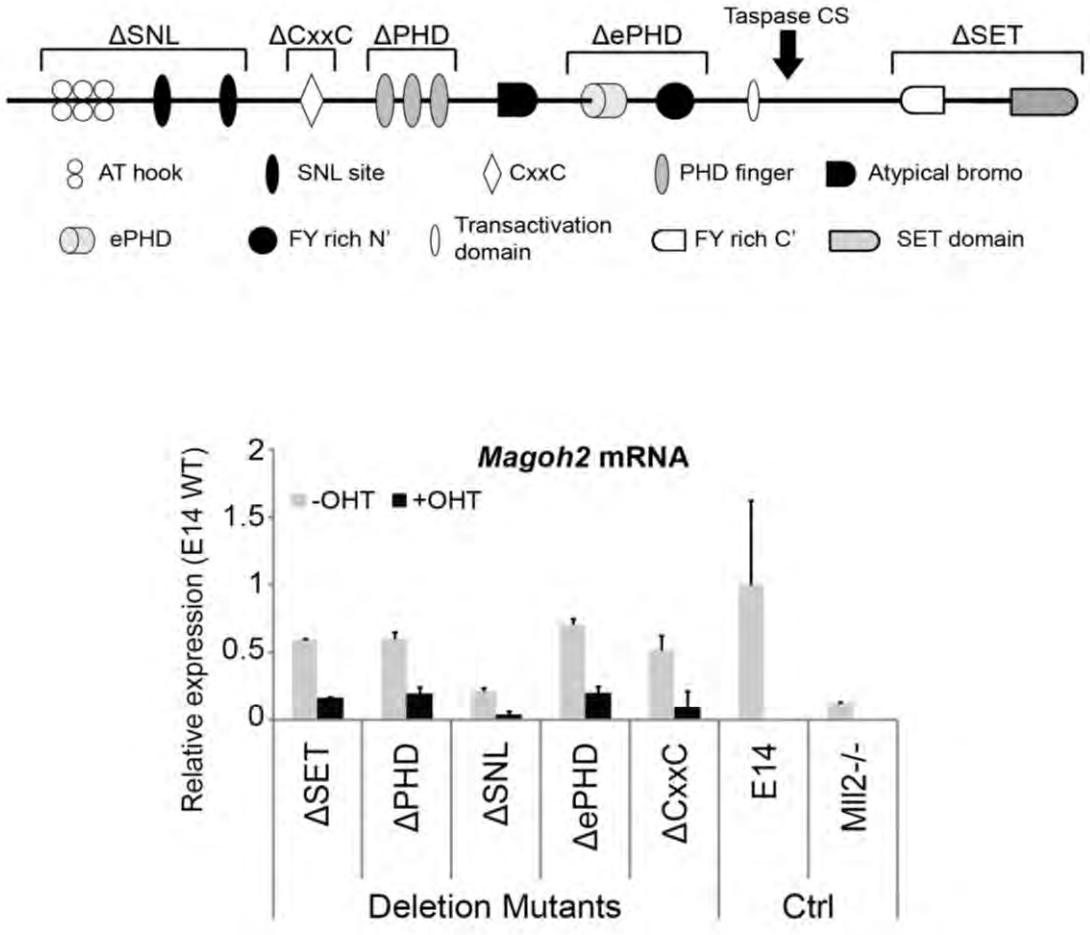

Figure 3.30. The complete MLL2 protein is required to maintain *Magoh2* expression. Top panel: *Mll2*$^{F/F}$ cells were transfected with *Mll2* BACs carrying mutants of MLL2 where specific domains were deleted. The schematic illustrates the deleted domains in each of the cell lines used. **Lower panel:** *Magoh2* expression analysis in *Mll2*$^{F/F}$ cells harbouring the mutant BACs. Bars represent the mean +sd of at least 4 measurements (2 biological x 2 technical replicates)

3.4. MLL2 in haemopoiesis

MLL1 is implicated in both normal haemopoiesis and the establishment of haematological malignancies. The role of MLL2 in this process is unclear as *Mll2* knock-out mouse embryos die early in development due to general tissue disorganisation and widespread apoptosis (Glaser et al., 2006). There are, however, indications that MLL2 is involved in macrophage function (Austenaa et al., 2012). We decided to study the involvement of MLL2 in early haemopoiesis to find out whether there was a requirement for MLL2 in early blood development that may be concealed by the early embryonic lethality observed by Glaser et al. (2006). To this end, we differentiated $Mll2^{F/F}$ ES cells towards blood lineages and examined the consequences of *Mll2* deletion.

3.4.1. Production of macrophages from embryoid bodies is abolished in the absence of *Mll2*

We first examined the effects of *Mll2* deletion on macrophage differentiation by performing a macrophage release assay using $Mll2^{F/F}$ and $Mll2^{Fc/Fc}$ cells. In these assays, cells were cultured in semi-solid media containing the cytokines M-CSF and IL3 which instruct macrophage differentiation. The cells initially form embryoid bodies (EBs) which later start releasing macrophages seen as a halo of round highly refractive cells around the EBs. Our first observation was that $Mll2^{Fc/Fc}$ cells had a strongly reduced EB forming potential as compared to $Mll2^{F/F}$ cells (*Figure 3.31 Top panel*). The cells were monitored for 19 days. In that time, a proportion of the $Mll2^{F/F}$ derived EBs generated macrophages unlike the $Mll2^{Fc/Fc}$ where no macrophage production was detected (*Figure 3.31 Lower panel*). However, both the $Mll2^{F/F}$ and $Mll2^{Fc/Fc}$ cells exhibited much lower EB forming potential and macrophage production than E14 wild type cells (*data not shown*), making comparisons between different cellular states difficult. To resolve this issue, $Mll2^{F/F}$ cells were subcloned and selected for increased EB forming potential and macrophage release. Two subclones (sc9 and 11) were selected to repeat the macrophage release assay (*Figure 3.32*). Almost all EBs derived from these subclones were surrounded by a macrophage halo after 15 days in the semi-solid culture medium. Contrastingly, when *Mll2* was deleted (8 days) before the cells were transferred in the semi-solid medium, macrophage production was

completely abolished. These results demonstrate a clear requirement for MLL2 in macrophage differentiation.

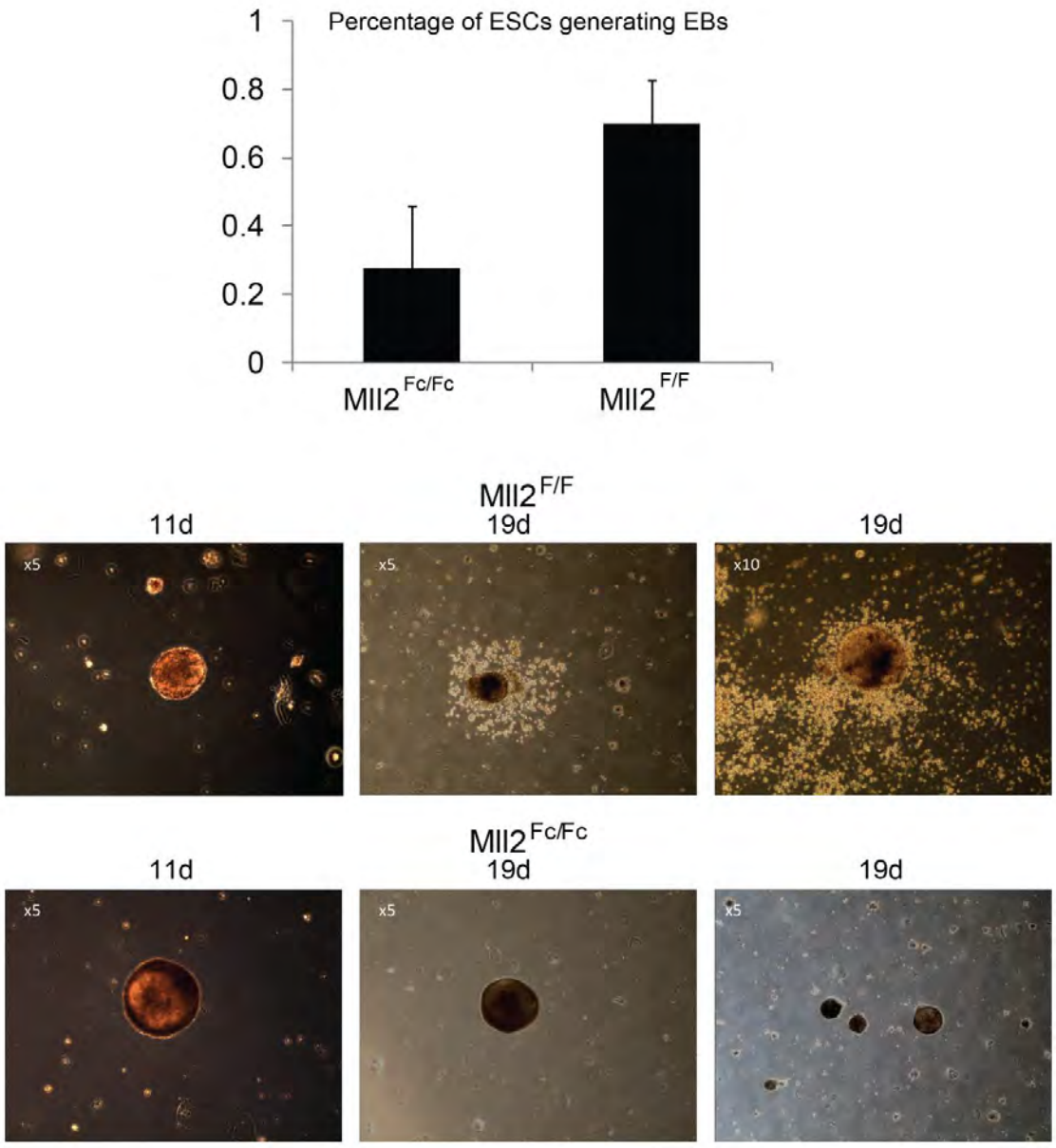

Figure 3.31. MLL2 depletion results in reduced embryoid body formation and macrophage production. Top panel: *Mll2*$^{F/F}$ and *Mll2*$^{Fc/Fc}$ cells were cultured in semi-solid media to generate embryoid bodies. Embryoid body formation capacity is calculated by counting the embryoid bodies formed and expressing them as a percentage of the number of cells initially seeded. The bars represent the mean +sd of embryoid bodies formed in 3 separate experiments as a percentage of the total number of cells seeded. Over 100 embryoid bodies were counted in each of the 3 experiments.**. Lower panel:** *Mll2*$^{F/F}$ (top row) and *Mll2*$^{Fc/Fc}$ (bottom row) cells were cultured in semi-solid media to generate embryoid bodies that would further differentiate and produce macrophages. Over 500 embryoid bodies were examined for each of the cell lines. None of the *Mll2*$^{Fc/Fc}$ derived embryoid bodies produced macrophages.

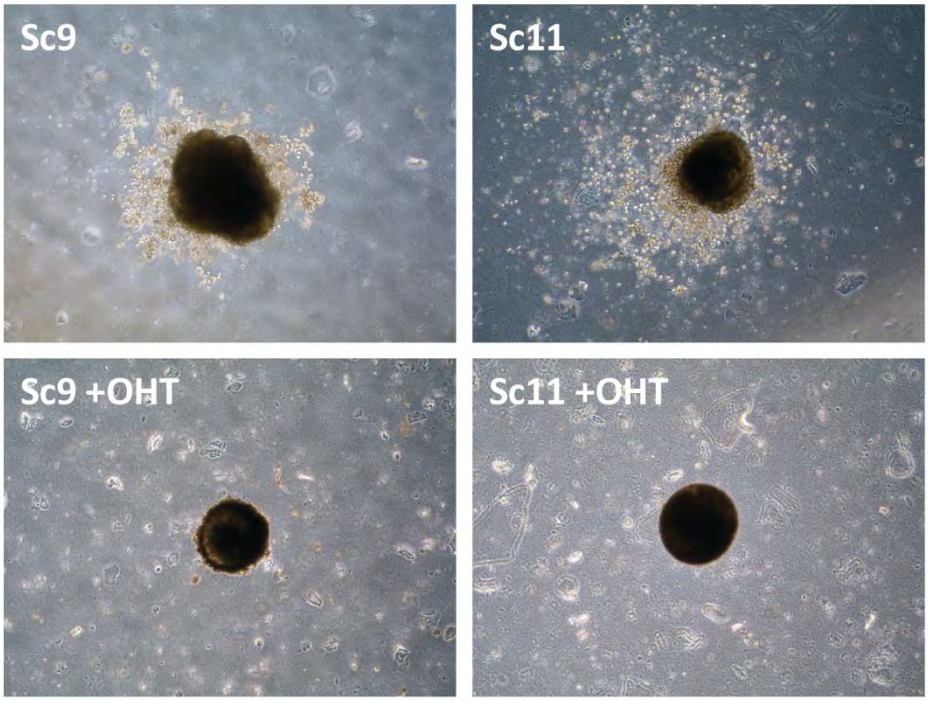

Figure 3.32. Verification of the observed differentiation phenotype using subcloned lines. The *Mll2*$^{F/F}$ cells were sub-cloned and re-subjected to the macrophage release assay. Two individual sub-clones (sc9 and sc11) were employed. 300 embryoid bodies were examined for each condition.

3.4.2. MLL2 is required for the specification of haemangioblast and haemopoietic progenitor cells

Macrophages differentiate via a well-described series of intermediate myeloid progenitor stages (reviewed in Chow et al., 2011). The experiments described in 3.4.1 did not reveal at which stage macrophage differentiation was blocked and the study by Austenaa et al. (2012) demonstrated that macrophage differentiation from haemopoietic stem cells in adult mice is not dependent on MLL2. These results suggest that the defect is very early in haemopoietic development, possibly before the

establishment of the haemopoietic system. To identify when exactly MLL2 is required during haemopoietic differentiation, we used a differentiation system that allowed us to delete *Mll2* at discrete developmental stages (Fehling et al., 2003; Lancrin et al., 2009; Pearson et al., 2008) and monitor how this loss affects the generation of haemopoietic progenitor cells (*Figure 3.33*). Cells were allowed to form EBs in suspension by culturing in low adherence bacteriological Petri dishes for approximately 3 days. FLK1 (VEGF receptor) expressing haemangioblast cells were enriched by MACS and cultured for another 3 days in the presence of VEGF and IL6. During this time the haemangioblast cells generate the so-called haemogenic endothelium – a tissue that expresses both endothelial and haemopoietic markers (TIE2, C-KIT, CD41). Haemogenic endothelium cells finally down-regulate the endothelial marker TIE2 and undergo a morphological transition becoming free-floating haemopoietic progenitors, which still express C-KIT and CD41.

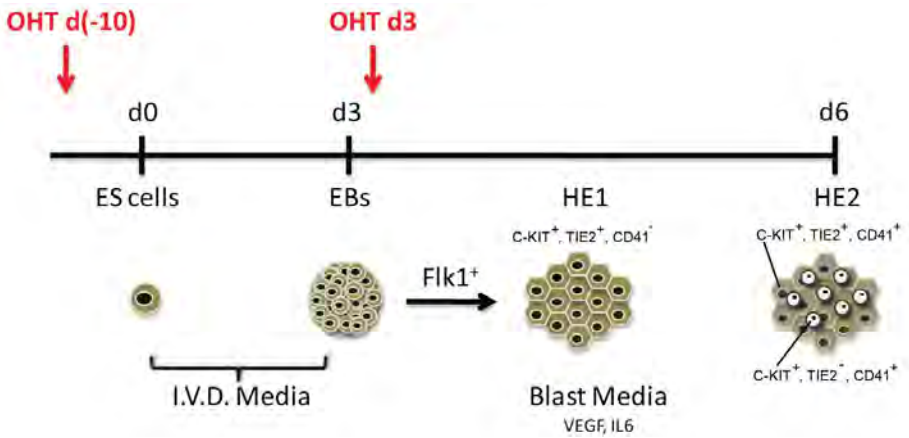

Figure 3.33. Schematic of the *in vitro* differentiation system. *Mll2*$^{F/F}$ cells were cultured in suspension for 3 days without LIF to form EBs. *Flk1* expressing haemangioblast enriched cells were purified by MACS. The purified cells were cultured in the presence of VEGF and IL-6 to generate haemogenic endothelium (hexagonal cells in the schematic) and haemopoietic progenitor cells (round white cells in the schematic). *Mll2* was deleted either in ES cells, 10 days prior to LIF withdrawal or in the *Flk1* positive haemangioblast enriched cells at day 3.

Mll2$^{F/F}$ cells were subjected to this differentiation method and approximately 35% of the cells were found to be expressing FLK1 after 3 days (*Figure 3.34*). When *Mll2* was deleted in the ES cell stage, FLK1 expression was completely abolished, revealing a requirement for MLL2 at this very early developmental stage. The inability of the MLL2-lacking cells to initiate *Flk1* expression during differentiation results in complete loss of haemangioblast cells which could account for the total lack of macrophage production from *Mll2*$^{Fc/Fc}$ derived EBs described earlier (*Figures 3.31, 3.32*).

To obtain a first insight into the reasons of the absence of *Flk1* expression we obtained unpublished ChIP-sequencing data generated in the A.F. Stewart and H.G. Stunnenberg labs. These assays examined H3K4me$_3$ distribution in E14 WT and *Mll2*$^{-/-}$ ES cells and MLL2 distribution in E14 WT ES cells. Interestingly, the data show that *Flk1* is an MLL2 target in ES cells (*Figure 3.35*). Unlike *Magoh2* however, H3K4 trimethylation was unaffected on the *Flk1* CpG island promoter in *Mll2*$^{-/-}$ ES cells, suggesting that – at least at the ES cell stage – loss of MLL2 can be compensated for by another SET1/MLL family member.

We then deleted *Mll2* in FLK1 positive haemangioblast cells (post MACS enrichment) to examine whether MLL2 is required at later developmental stages as well. The cells were harvested on day 6 of differentiation and subjected to FACS analysis for C-KIT, TIE2 and CD41 expression (*Figure 3.36*). These three surface markers define three distinct cell populations: haemogenic endothelium I (C-KIT$^+$, TIE2$^+$, CD41$^-$),

haemogenic endothelium II (C-KIT$^+$, TIE2$^+$, CD41$^+$) and haemopoietic progenitors (C-KIT$^+$, TIE2$^-$, CD41$^+$). Although the percentage of C-KIT positive cells in induced and control populations is similar, the pattern is strikingly different. Non-induced cells exhibit high levels of C-KIT staining while cells lacking MLL2 did not express C-KIT to the same extent. This suggests that MLL2 is also required after haemangioblast specification to correctly upregulate *c-kit* expression. Intriguingly, the *c-kit* CpG island promoter is also an MLL2 target in ES cells (*data not shown, unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*).

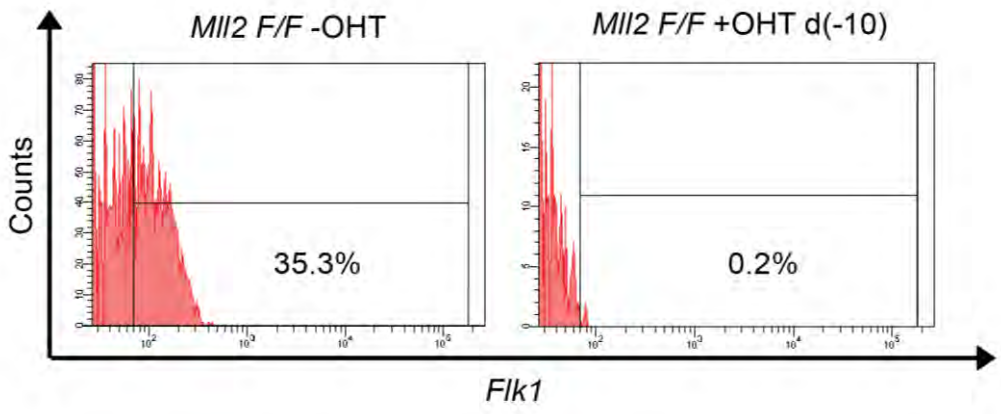

Figure 3.34. ES cells lacking MLL2 cannot differentiate into FLK1 expressing haemangioblasts. Differentiation of *Mll2*$^{F/F}$ ES cells ±OHT into embryoid bodies, followed by FACS analysis of FLK1 expression on day 3.

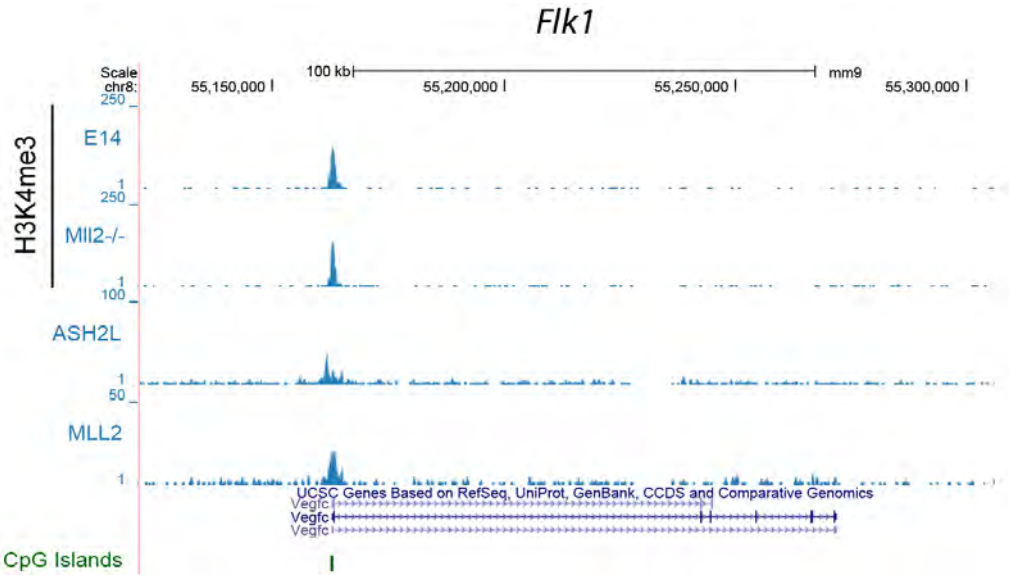

Figure 3.35. *Flk1* is a direct MLL2 target in ES cells (*unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). H3K4 trimethylation on the *Flk1* promoter in ES cells does not depend on MLL2, however, correct upregulation of *Flk1* during ES cell differentiation requires the presence of MLL2 as shown in *Figure 3.34.*

When C-KIT positive cells were examined further for the presence of TIE2 and CD41, we observed only minor differences suggesting that the presence of MLL2 is not critical at this stage. Intriguing as these results may be, we were unable to reproduce them due to technical issues affecting culture conditions.

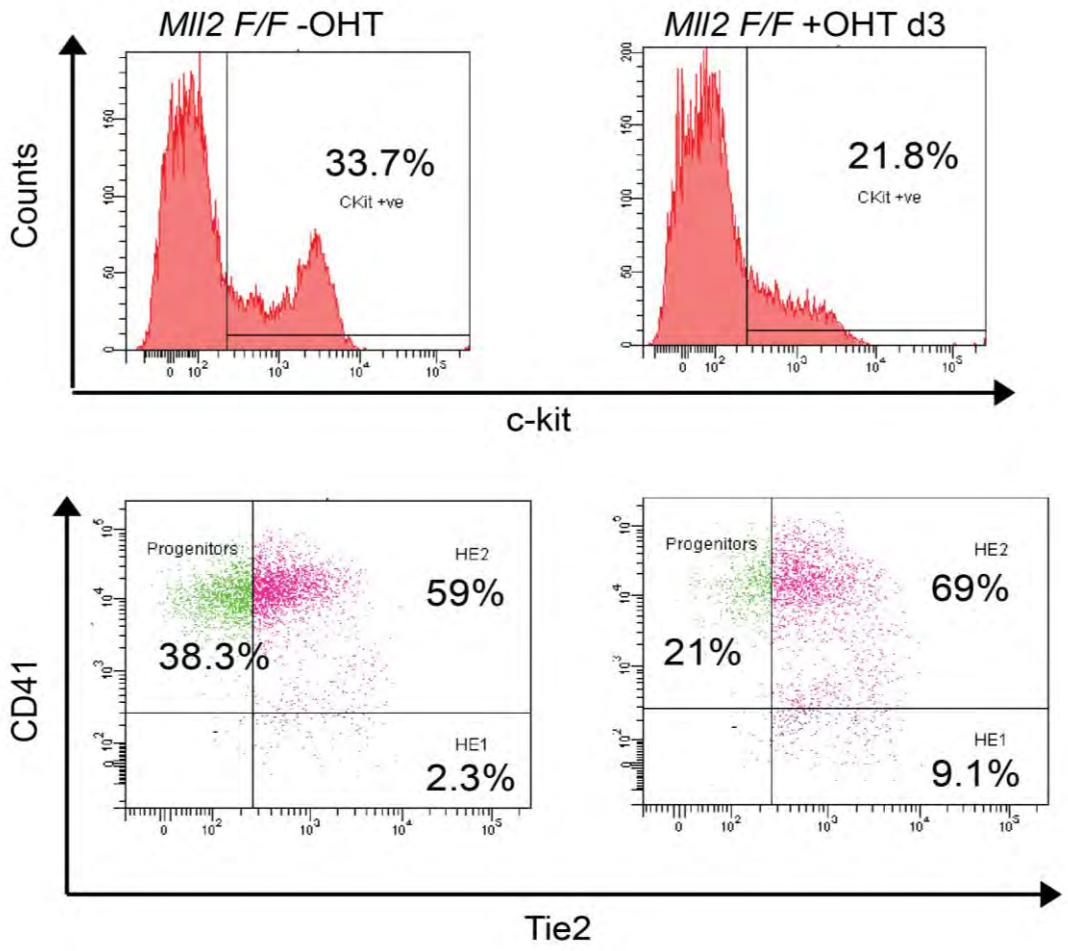

Figure 3.36. Cells lacking MLL2 fail to correctly upregulate *c-kit* and exhibit a defect in progression through the different haemogenic endothelium stages to haemopoietic progenitors. The blast cultures were harvested on day 6 and analysed by fluorescence activated cell sorting. The C-KIT positive cells in the top graphs were gated and further analysed for CD41 and TIE2 expression in the bottom graphs.

3.4.3. MLL2 is required for the correct establishment of transcriptional programs during differentiation

Having established that MLL2 was vital for normal haemopoiesis, we further explored the mechanism behind this phenomenon. We therefore tested whether this differentiation defect was solely due to the observed *Flk1* deficiency or whether there were more factors depending on MLL2 for correct expression. To address this question, we measured the mRNA expression levels of genes encoding key developmental factors over a differentiation time-course. Haemangioblast enriched FLK1 positive cells were grown as described above with or without OHT and the resulting populations were harvested daily over the next 3 days (*Figure 3.37 Top panel*). Of the genes examined, *Scl/Tal1*, *Gata2, Gfi1* and *Runx1* appeared to be regulated normally when *Mll2* was deleted. *Pu.1* exhibited a delayed induction pattern which resulted in a delayed induction of *Csf1R* one of the best characterised PU.1 targets (DeKoter et al., 1998; Himes et al., 2001; Krysinska et al., 2007).

Importantly, MLL2 was found on a CpG island overlapping the *Pu.1* terminator region in ES cells (*Figure 3.38, unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). This CpG island is marked by H3K4me$_3$, a histone mark almost exclusively associated with promoters (Mikkelsen et al., 2007). Deposition of the H3K4me$_3$ mark on that region is dependent on the presence of MLL2 in ES cells. However, the functional relevance of this early chromatin marking is unknown, as *Pu.1* expression

is only observed in haemopoietic progenitor cells (Lichtinger et al., 2012, *in press* and Cheng et al., 1996). It is therefore equally possible that decreased *Pu.1* expression is a result of reduced numbers of progenitor cells in the *Mll2*$^{F/F}$ +OHT blast cultures (*Figure 3.36*). Further investigation is deemed necessary to elucidate the precise role of MLL2 in haemopoietic differentiation.

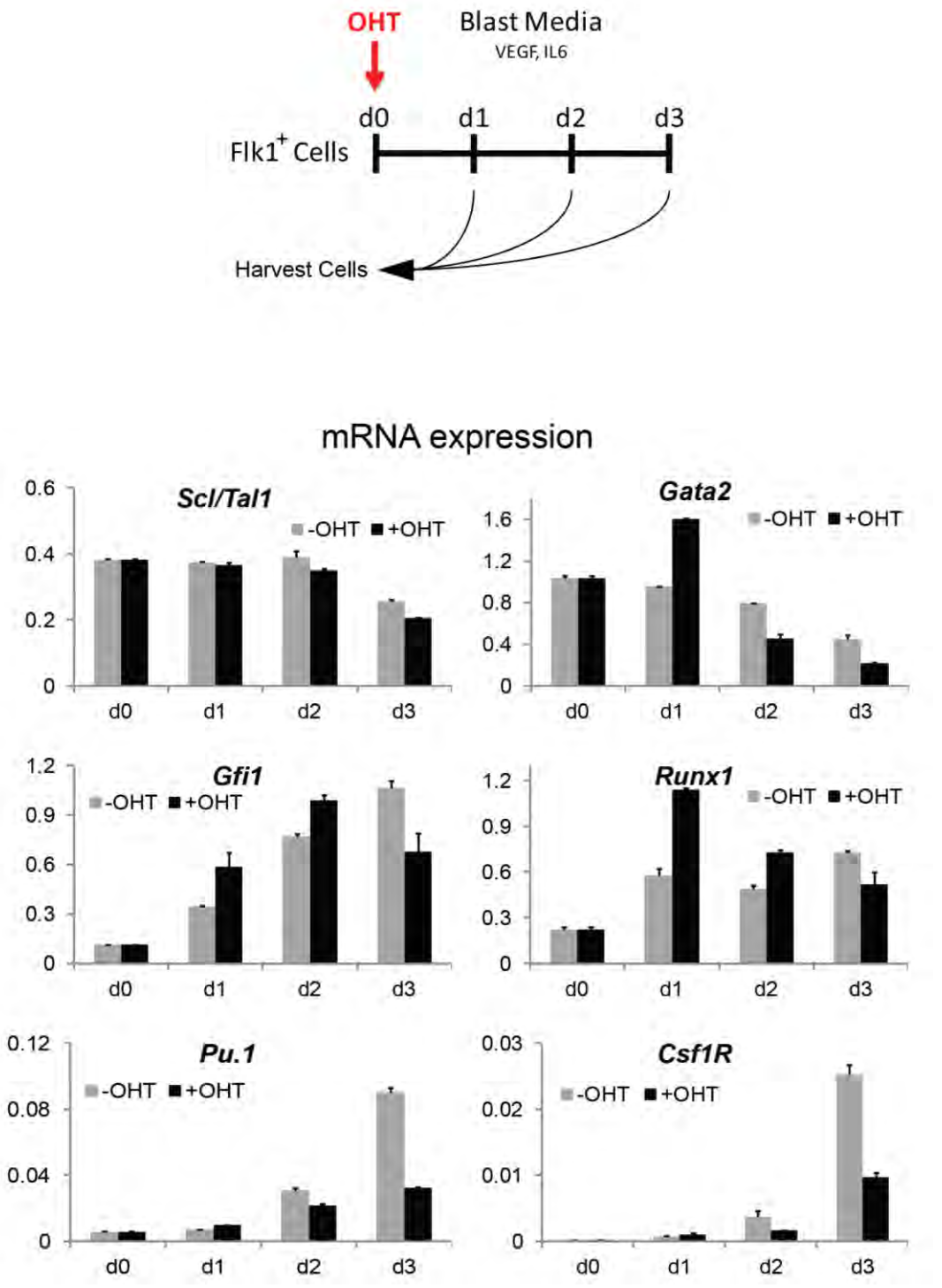

Figure 3.37. Deletion of *Mll2* perturbs the normal gene expression pattern in differentiating cells. Top panel: *Mll2*$^{F/F}$ sc9 cells were differentiated using the *in vitro* differentiation system described in paragraph 2.2.3. *Mll2* was deleted by addition of OHT in the culture medium, immediately after purification of the haemangioblast cells. The resulting differentiating cells were harvested over the next 3 days and used for expression analysis. **Bottom panel:** Expression levels of genes involved in differentiation of haemangioblasts to haemopoietic progenitors via haemogenic endothelium were measured by RT-QPCR. The bars represent the mean +sd of 2 technical replicates.

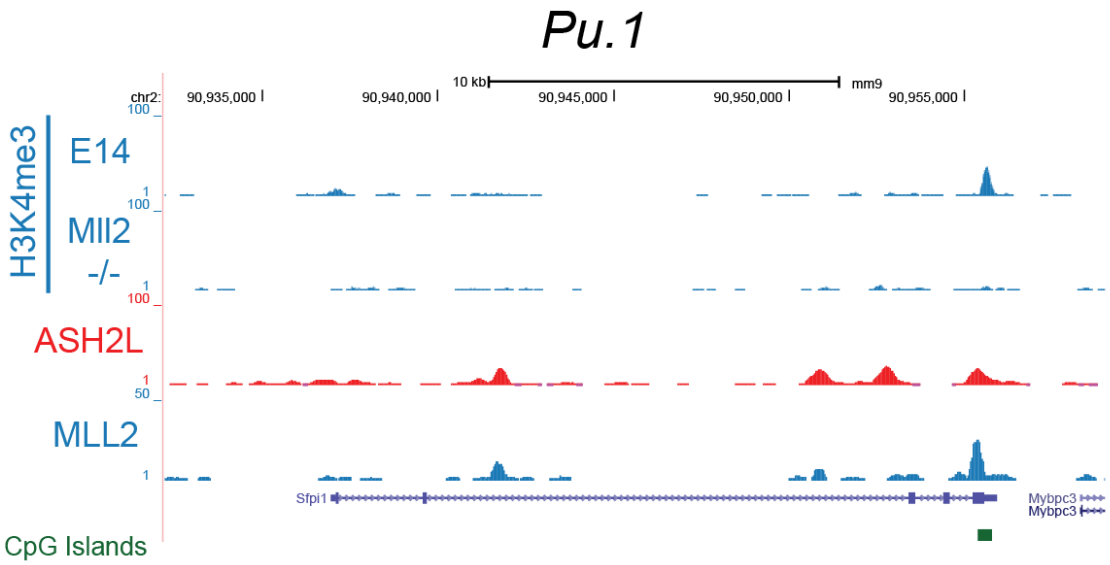

Figure 3.38. The *Pu.1* terminator is a direct MLL2 target in ES cells (*unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). H3K4 trimethylation on the *Pu.1* terminator in ES cells depends on the presence of MLL2.

4. DISCUSSION

Set1 is the only H3K4 methyltransferase in yeast and has been shown to be recruited to the initiating RNA polymerase II (Ng et al., 2003) via interaction with the Paf1 complex (Krogan et al., 2003). Therefore, in yeast trimethylation of lysine 4 on histone H3 occurs after the basal transcription machinery has assembled and the RNA polymerase II has initiated transcription. Additionally, deletion of Set1 in yeast resulted in complete loss of H3K4me_3 but only exhibit small growth defects (Briggs et al., 2001). Several studies in mammalian cells have reported that depletion of the H3K4 methyltransferase MLL1 results in abnormal transcriptional silencing, RNA polymerase II depletion and DNA methylation of a subset of the clustered homeobox genes (Milne et al., 2005a; Terranova et al., 2006; Wang et al., 2009) and death in the early embryonic stages. These apparently contradicting results could either point to a fundamental difference between yeast and higher organisms or point to a cooperative mechanism that requires the synergistic assembly of multiple components. We therefore exploited the exquisite dependency of *Magoh2* on MLL2 to gain mechanistic insights into how MLL proteins are involved in the activation and maintenance of transcriptional activity of higher eukaryotic genes.

4.1. MLL2 in transcriptional regulation

4.1.1. MLL2 is required for the formation of *Magoh2* open chromatin and expression

It has previously been shown that *Magoh2* is transcriptionally silenced in *Mll2*$^{-/-}$ ES cells (Glaser et al., 2009). Here we extended these analyses by showing that this gene's promoter resides in densely packaged chromatin in *Mll2*$^{-/-}$ ES cells. Additionally, H3K9ac, H3K4me$_3$ and the basal transcription machinery are absent from the *Magoh2* promoter, which carries a high level of DNA methylation in these cells. From our experiments, it appears that MLL2 on *Magoh2* functions in a manner similar to MLL1 on the homeobox gene clusters (Milne et al., 2005b; Terranova et al., 2006; Wang et al., 2009). Taken together our and previously published results suggest that MLL proteins have a crucial role for transcription and may be required for assembly of the RNA polymerase II pre-initiation complex, recruitment of HATs to active promoters and protection from DNA methylation.

4.1.2. MLL2 is required for the binding of RNA-polymerase II at *Magoh2*

The inducible *Mll2* knock-out system allowed us to establish the order of events leading to *Magoh2* silencing and demonstrate that *Magoh2* silencing occurs in a sequential fashion. After MLL2 depletion, the levels of the H3K4me$_3$ histone mark decrease (herein and Glaser et al., 2009) with a concomitant decrease in both H3K9 acetylation and RNA polymerase II association with the *Magoh2* promoter. In contrast to our findings, a recent study (Clouaire et al., 2012) has deleted *Cfp1*, a key component of the SET1A and SET1B complexes. The data presented suggest that loss of the H3K4me3 mark is inconsequential for gene transcription. These seemingly contradictory results can be reconciled if what we have hypothesised earlier is true: mammalian SET1A and B operate similarly to yeast Set1 and are fundamentally different to mammalian MLL proteins.

Our rescue experiments show that MLL2 catalytic activity is required for transcriptional maintenance of *Magoh2*, as a ΔSET *Mll2* BAC transgene could not maintain *Magoh2* expression. The same experiments show that other domains of MLL2 are also required to maintain *Magoh2* expression. We speculate that the other domains may be important for MLL2 recruitment to *Magoh2* or interactions with accessory factors. By extension, it is possible that disruption of such domains on leukaemogenic MLL1 fusion proteins will block their function, making them potential targets for small molecule therapeutic agents. The CxxC domain appears

to be a good candidate target for drug development as it is required for their leukaemogenic potential (Ayton et al., 2004) and differences in this domain preclude MLL2 fusion mediated leukaemogenesis (Bach et al., 2009).

Loss of RNA polymerase II results in *Magoh2* transcriptional silencing and possibly nucleosome repositioning over the *Magoh2* promoter, as deduced from MNase footprinting experiments. Such an intricate competition for promoter binding between RNA polymerase II and nucleosomes has been reported previously in both humans and yeast (Lomvardas and Thanos, 2001; Schwabish and Struhl, 2004).

In summary, from our data it is likely that MLL2 forms a platform that coordinates the synergistic assembly of the basal transcription machinery via multiple protein-protein interactions. Interestingly, the *trans*-activation domain of MLL1 has been shown to interact directly with CBP/p300-containing HAT complexes (reviewed in Cosgrove and Patel, 2010). HAT complexes are known to act as large multi-protein scaffolds that can mediate interactions between TFs and the basal transcription machinery (reviewed in Chan and La Thangue, 2001). In this context, MLL2 may be the factor that recruits CBP/p300-containing complexes – via its *trans*-activation domain – which in turn recruit the transcriptional apparatus. This may be assisted by $TF_{II}D$ binding to H3K4 trimethylated nucleosomes (van Ingen et al., 2008; Vermeulen et al., 2007) flanking the *Magoh2* promoter. Consistent with our results on the *Magoh2* promoter,

this model predicts that loss of MLL2 would result in loss of histone acetylation and RNA polymerase II.

4.1.3. H3K4 trimethylation on the *Magoh2* promoter does not depend on the presence of RNA polymerase II

Our experiments using transcriptional blockers suggest that in the presence of MLL2, RNA polymerase II and transcriptional elongation are largely dispensable for the maintenance of H3K4 trimethylation. This result is consistent with the hypothesis that MLL proteins operate differently to yeast Set1 and may have a causal role in transcriptional activation. Previous studies have hinted to such a mechanism as depletion of MLL1 resulted in changes in RNA polymerase II distribution on MLL1 target genes (Milne et al., 2005a; Wang et al., 2009). Additionally, it has been shown that CpG islands recruit H3K4 methyltransferases regardless of their transcriptional activity (Thomson et al., 2010). Our experiments support the idea that neither active transcription nor the presence of a RNA polymerase II pre-initiation complex are required to maintain the H3K4me$_3$ mark over an active CpG island promoter, further suggesting that MLL proteins act upstream of RNA polymerase II recruitment. Our experiments also suggest that loss of transcription is not sufficient to induce removal of the H3K4me$_3$ mark from the *Magoh2* promoter. Compelling as these results may be, there is a possibility that there are no

functional H3K4 demethylases in the cells, due to global loss of transcription. This would also explain the increase in H3K4me$_3$ levels we observed on the *Magoh2* promoter following α-amanitin treatment. If our results hold true, they support the idea that histone modifications have a role in transcriptional regulation – especially since a ΔSET MLL2 cannot exert maintain *Magoh2* expression – and are not merely by-products of active transcription. Indeed, another research group (Fenouil et al., 2012) had very similar findings after depletion of RNA polymerase II on a genome-wide scale (Dr. P. Cauchy, personal communication). Specifically, they observed that after α-amanitin treatment, CpG island promoters maintain the H3K4me$_3$ mark while this mark is lost on most non-CpG island promoters. Their results suggest that there are still functional H3K4 demethylases in the cells for at least 36 hours of α-amanitin treatment. Additionally, these results are consistent with our hypothesis that H3K4me$_3$ is independent of transcription, at least on CpG island promoters.

4.1.4. Loss of MLL2 from the *Magoh2* promoter leads to rapid DNA methylation

How CpG islands are protected from DNA methylation has been a long-standing subject of research. Our data from the inducible *Mll2* knock-out system clearly demonstrate that loss of the H3K4me$_3$ mark results in the

rapid onset of DNA methylation at multiple CpG dinucleotides. Moreover, transcriptional elongation and RNA polymerase II appear to be dispensable with regards to protection from DNA methylation, as illustrated by our experiments using DRB and α-amanitin.

A previously proposed model (Turker, 2002) suggests that the methylation status of a genomic region is controlled by antagonistic methylation-promoting and methylation-blocking factors. Active promoters were thought to act as roadblocks to the expansion of DNA methylation simply by virtue of being transcribed. Supporting this idea, a recent study (Hitchins et al., 2011) demonstrated that decreased *MLH1* expression – due a single nucleotide polymorphism at a HSF binding site– resulted in increased DNA methylation of the *MLH1* promoter. Conversely, a mono-allelic insertion of a SP-family binding site in the human *RIL* promoter was shown to increase methylation protection of that particular allele without affecting *RIL* expression levels (Boumber et al., 2008). It has been demonstrated that H3K4me$_3$ may confer protection from DNA methylation (Ooi et al., 2007) as the DNA methyltransferase common subunit – DNMT3L – can only interact with unmethylated H3K4. As such MLL catalytic activity would preclude DNA methyltransferase recruitment, thus conferring protection from DNA methylation to MLL target promoters. Unfortunately, neither of the two aforementioned studies (Boumber et al., 2008; Hitchins et al., 2011) measured H3K4me$_3$ levels over the respective promoters. It is possible that these mutations affect the recruitment of MLL proteins and H3K4me$_3$ levels. Our experiments on the *Magoh2* promoter

could explain the above observations as we have shown that loss of the H3K4me$_3$ mark can result in silencing and DNA methylation.

The original model proposed by Turker may very well hold true for SET1 regulated promoters, as at those promoters H3K4me$_3$ deposition may be dependent on transcription initiation (much like in yeast). In that case, transcriptional activity would be required for protection from DNA methylation either directly or indirectly via the recruitment of SET1A/B to the initiating RNA polymerase II and H3K4 methylation. Our data suggest that MLL-regulated promoters are protected from DNA methylation due to the presence of the H3K4me$_3$ mark rather than their transcriptional status.

Glaser et al. (2009) reported that there is a significant increase in PcG mediated H3K27 methylation on the *Magoh2* promoter after *Mll2* deletion. The PRC2 catalytic subunit, EZH2, responsible for depositing this mark has been shown to directly associate with DNA methyltransferases (Vire et al., 2006), providing further insight into the mechanism of DNA methyltransferase recruitment to the *Magoh2* promoter after the H3K4me$_3$ mark has been removed. Importantly, this model may be applicable to other MLL-regulated CpG island promoters and may have implications for human disease. Indeed *Mll3* deletions and *de novo Mll4* loss-of-function mutations have been found with increased frequencies in patients with myeloid leukaemias and Kabuki syndrome respectively (Ng et al., 2010; Ruault et al., 2002). It remains to be seen which MLL target genes are deregulated and/or epigenetically silenced in patients with such mutations.

4.1.5. Re-expression of MLL2 can reactivate *Magoh2* after epigenetic silencing

DNA methylation is considered to be a stable epigenetic silencing modification (reviewed in Suzuki and Bird, 2008) and currently a number of models have been put forward to explain how this modification is removed. DNA methylation can be passively removed by dilution over consecutive DNA replication cycles with concomitant inhibition of the maintenance methyltransferase DNMT1 (reviewed in Wu and Zhang, 2010). Alternatively, DNA methylation can be erased enzymaticaly, although the mechanistic details are not entirely clear. The first proposed model suggested that the methyl-cytidine base is removed by the TDG and/or MBD4 glycosylases, creating an abasic site (Lucey et al., 2005; Petronzelli et al., 2000; Yoon et al., 2003b) which is recognised by the base excision repair machinery (reviewed in Wu and Zhang, 2010). Alternatively, it is more likely that methyl-cytosine is first deaminated into thymidine, generating a T:G mismatch which is the native TDG substrate (reviewed in Wu and Zhang, 2010). Additionally, it has been found that methyl-cytosine can be further modified by TET family proteins into hydroxymethyl-cytosine (Tahiliani et al., 2009), which can also be deaminated and/or processed by TDG/MBD4 (Hashimoto et al., 2012; Lucey et al., 2005; Petronzelli et al., 2000; Yoon et al., 2003b).

Although the mechanism for DNA demethylation is not entirely clear there is evidence for active enzymatic erasure of DNA methylation. Previous studies have reported removal of DNA methylation from the *IL-2* promoter in response to T-cell activation in as little as 20 minutes post stimulation (Bruniquel and Schwartz, 2003) and the *BDNF* promoter upon KCl depolarisation-mediated neuronal stimulation. Moreover, two recent studies reported cyclical methylation and demethylation of active promoters between individual rounds of transcription (Kangaspeska et al., 2008; Metivier et al., 2008), suggesting that methyl-moieties can be rapidly deposited and removed from DNA. Consistent with the proposed mechanisms for DNA demethylation they were able to show cyclical recruitment of DNA repair associated proteins to those promoters.

In this study we demonstrated that simple re-expression of *Mll2* is sufficient to overcome the repressive barrier posed by DNA methylation and reactivate *Magoh2* transcription, as illustrated by reactivation of the endogenous *Mll2* alleles in *Mll2*$^{-/-}$ ES cells. We have no indication as to which factors recruit MLL2 to the *Magoh2* promoter but these factors are able to recruit MLL2 regardless of the fact that the *Magoh2* promoter has become DNA methylated and resides in compacted chromatin.

Based on our results we propose a model for the role of MLL2 in *Magoh2* transcriptional regulation (*Figure 4.1*). After MLL2 depletion the H3K4me$_3$ mark is either actively removed or diluted out over consecutive cell divisions concomitantly with a loss of H3K9 acetylation. RNA polymerase II can no longer associate with the *Magoh2* promoter, as its

recruitment may be facilitated or stabilised by direct interactions of $TF_{II}D$ with the H3K4me$_3$ mark (van Ingen et al., 2008; Vermeulen et al., 2007) and interactions with CBP/p300 containing HAT complexes (reviewed in Chan and La Thangue, 2001; Cosgrove and Patel, 2010). Loss of the basal transcription machinery allows nucleosomes to be repositioned over the stretch of DNA previously occupied by the pre-initiation complex with a concomitant increase in DNA methylation.

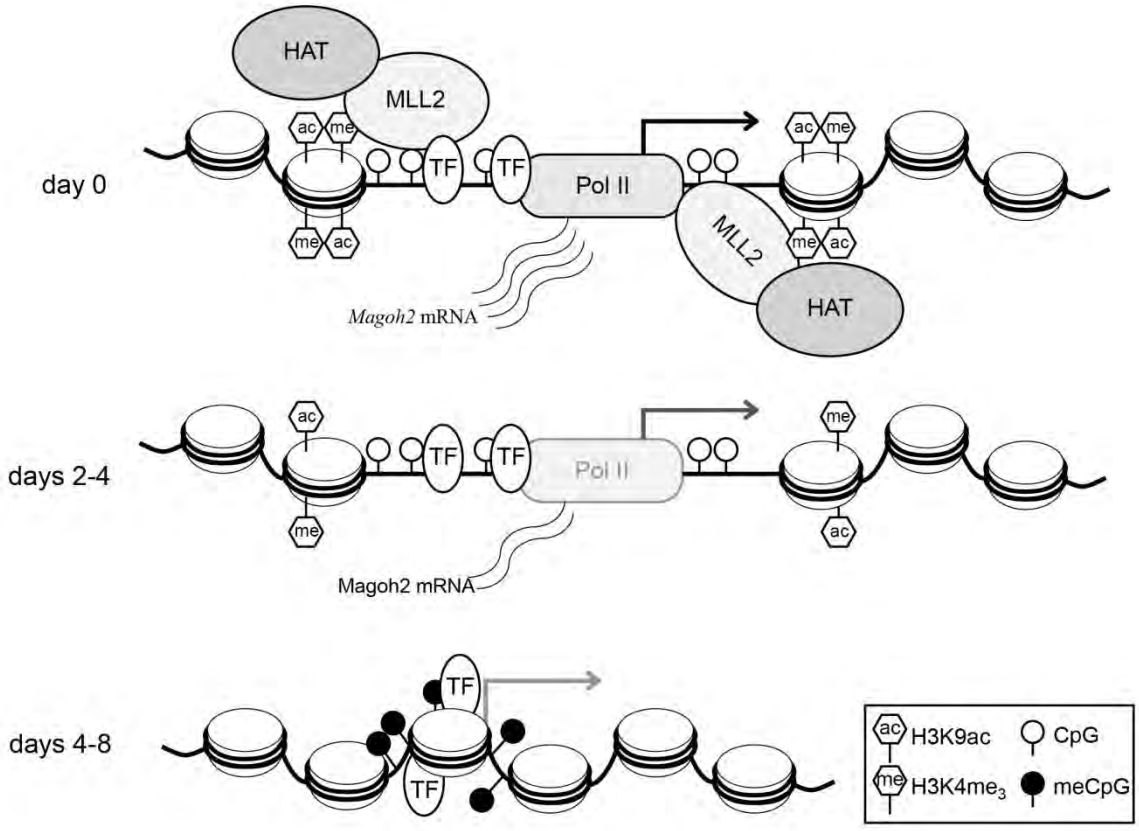

Figure 4.1. Model for *Magoh2* silencing following MLL2 depletion. MLL2 is normally recruited to the *Magoh2* promoter by TFs, interaction of the PHD fingers with H3K4me$_3$ or through CxxC domain binding to non-methylated CpG dinucleotides. After *Mll2* deletion, H3K4me3 and H3K9ac are depleted from the *Magoh2* promoter, resulting in eviction of RNA polymerase II, transcriptional silencing and nucleosome repositioning. Removal of the H3K4me$_3$ mark further results in DNA methylation and chromatin compaction.

4.1.6. *Mll2* in CpG island promoter regulation – future directions

Currently it is still unclear which protein factors recruit MLL2 to *Magoh2* and whether these are sequence-specific DNA binding proteins. Although we were unable to assign individual roles to SP1 and SP3 in the regulation of *Magoh2*, it is still possible that these factors compensate for each other's absence. It would therefore be important, to perform a careful molecular characterisation of the *Magoh2* promoter using point mutagenesis and reporter gene assays in ES cells in the presence and absence of MLL2, to precisely identify factor binding sites mediating the MLL2 effect. If the deletion of *Mll2* had no influence on transiently transfected templates, this could alternatively be investigated by BAC transgenesis using *Magoh2* constructs where individual binding sites were altered by recombineering.

Only *Magoh2* becomes transcriptionally silenced in response to *Mll2* deletion in ES cells. However, a number of other MLL2 target genes lose the H3K4me$_3$ mark from their promoters after *Mll2* deletion in ES cells (*unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). These genes are regulated by CpG island promoters and are bivalently marked and transcriptionally silent in ES cells. The role of H3K4me$_3$ on those promoters is therefore not entirely clear. Moreover, it is currently unknown whether loss of MLL2/H3K4me$_3$ from these promoters in ES cells affects transcriptional regulation of these genes later in development. To address this question we would like to perform genome wide MLL2,

$H3K4me_3$, $H3K27me_3$, DNA methylation and RNA expression analyses during *in vitro* differentiation towards blood and neuronal lineages. In this experimental setting we could delete *Mll2* at different developmental stages and examine the chromatin dynamics and transcriptional status of MLL2 regulated CpG island promoters. Such experiments would provide valuable insight into chromatin mediated gene-priming mechanisms and shed light on the dynamics of MLL2 distribution during development.

4.2. MLL proteins, the cell cycle and cell survival

There have been numerous reports linking MLL1 to the cell cycle. This function of MLL1 is mediated by its involvement in transcription of the CDK inhibitors p27Kip1 and p18Ink4c (Milne et al., 2005b) and several S-phase specific genes, through interactions with HCF-1 and E2F family proteins (Tyagi et al., 2007). Taspase 1 mediated cleavage of MLL1 was shown to be critical for MLL1 function in cell cycle progression (Takeda et al., 2006). Our data and those by Lubitz et al. (2007) show that MLL2, unlike MLL1 does not affect cell cycle progression but is partially required for ES cell survival, as illustrated by increased ES cell apoptosis in the absence of MLL2. According to Lubitz et al. (2007), *Mll2*$^{-/-}$ cells appear to become apoptotic during or immediately after cell division and this apoptotic phenotype correlates with decreased expression of the anti-apoptotic factor *Bcl2* in these cells.

4.3. MLL2 plays a role in haemopoiesis

4.3.1. MLL2 is required for macrophage differentiation and the generation of $FLK1^+$ cells from differentiating ES cells

MLL1 has been reported to play an important role in generation of haemopoietic stem cells and their potential to repopulate the bone marrow of irradiated mice (Ernst et al., 2004; McMahon et al., 2007). Austenaa et al. (2012) demonstrated that MLL2 is not required for macrophages differentiation in the bone marrow but is vital for their function. *Pigp* a CpG island promoter regulated gene is transcriptionally silenced in *Mll2*$^{-/-}$ macrophages. PIGP is a key enzyme in the GPI-anchor biosynthesis pathway (Kinoshita et al., 2008; Watanabe et al., 2000). PIGP depletion results in a complete loss of GPI anchored proteins and impaired macrophage responses to LPS, mediated by loss of GPI-anchored CD14 (Austenaa et al., 2012). The *Pigp* silencing mechanism appears to be similar to what we and Glaser et al. (2009) described for *Magoh2* (transcriptional silencing, loss of H3K4me$_3$ and acquisition of H3K27me$_3$).

So far no role of MLL2 in early haemopoiesis had been described, because *Mll2* knock-out embryos die early during development due to widespread apoptosis (Glaser et al., 2006) prior to the establishment of the haemopoietic system. Use of the *in vitro* ES cell differentiation system circumvented this problem and demonstrated that MLL2 is absolutely necessary for macrophage differentiation from EBs. Importantly, the

inducible *Mll2* knock-out system gives confidence that the observed differentiation defects are not due to clonal differences. Our results can be reconciled with the seemingly contradictory previously reported findings (Austenaa et al., 2012). Austenaa et al. (2012) used OHT-fed Cre-ERT2; $Mll2^{F/F}$ adult mice. In these mice the haemopoietic system has already been established before deletion of *Mll2.* Their data demonstrate that differentiation of macrophages from haemopoietic stem cells in the bone marrow is unaffected by *Mll2* deletion. However, it is possible that MLL2 is required for the establishment of haemopoietic stem and progenitor cells earlier in development.

To further investigate this hypothesis we employed a more sophisticated culture system that mimics embryonic haemopoiesis via generation of haemogenic endothelium (Fehling et al., 2003; Lancrin et al., 2009; Pearson et al., 2008).Our preliminary findings using this system suggest that MLL2 is involved in establishment of the haemopoietic system. Deletion of *Mll2* in ES cells resulted in complete ablation of the FLK1 positive population in EBs. A similar effect has been previously observed by *Flk1* expression analysis in $Mll2^{-/-}$ derived EBs (Lubitz et al., 2007). The FLK1 expressing population is enriched in haemangioblasts, the cells that give rise to haemogenic endothelium and haemopoietic progenitors. Depletion of this population could result in severe defects in both primitive and definitive haemopoiesis, as *Flk1* has been shown to be absolutely required for both endothelial and haemopoietic development (Shalaby et al., 1997).

Flk1 is a direct target of MLL2 in ES cells (*Figure 4.2, unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). However, deletion of *Mll2* did not affect H3K4me$_3$ levels on the *Flk1* CpG island promoter in ES cells. It is possible that *Flk1* is co-regulated by a different SET1/MLL family member in ES cells but as development progresses this redundancy collapses and MLL2 is then required for the maintenance of H3K4me$_3$ on the *Flk1* promoter. In the absence of MLL2, we would expect an aberrant silencing of *Flk1* by a mechanism similar to what we described previously for *Magoh2.* When we deleted *Mll2* in haemangioblast cells, we observed suboptimal up-regulation of C-KIT expression in progenitor cells derived from these cells, suggesting that MLL2 is required at this later stage as well. Although the *c-Kit* CpG island promoter is also a direct MLL2 target, no change in H3K4me$_3$ was observed in *Mll2*$^{-/-}$ ES cells (*unpublished data from the A.F. Stewart and H.G. Stunnenberg labs*). Future experiments using purified haemangioblast cells should clarify the histone modification status of these genes in the absence of MLL2.

4.3.2. MLL2 is required for the correct timing of expression of myeloid genes

Examination of the expression kinetics of genes involved in haemopoietic differentiation showed that the timing of expression of such genes appears to be perturbed by MLL2 depletion. *Scl/Tal1*, *Gata2* and *Runx1* are direct MLL2 targets in ES cells and *Mll2* deletion results in depletion of the H3K4me$_3$ mark on their promoters (*not shown, unpublished ChIP-seq data from the A.F. Stewart and H.G. Stunnenberg labs*). However, these genes appear to be regulated normally during blast culture, suggesting that in these cases robust regulatory mechanisms are in effect which can overcome H3K4me$_3$ depletion on those promoters. *Pu.1* (*Sfpi1*) is another direct MLL2 target in ES cells (*Figure 3.38*), although no RNA is expressed from its main promoter in these cells (Hoogenkamp et al., 2009). MLL2 is recruited to the *Pu.1* terminator, which overlaps a CpG island and *Mll2* deletion results in loss of the H3K4me$_3$ mark from the *Pu.1* terminator in ES cells. It is unclear whether H3K4me$_3$ depletion on the terminator is of functional significance for subsequent *Pu.1* activation. Our blast culture FACS analyses demonstrated decreased numbers of haemopoietic progenitors in the absence of MLL2 (*Figure 3.36*). As *Pu.1* is only upregulated in these cells (Lichtinger et al., 2012, *in press* and Cheng et al., 1996), it is possible that the abnormal kinetics we observed are due to an overall depletion of this population rather than a real transcriptional defect. Further

experimentation is required to clarify this issue. If *Pu.1* expression is indeed delayed, it will cause delayed expression of *Csf1R* as this gene is absolutely dependent on the presence of PU.1 (DeKoter et al., 1998; Himes et al., 2001; Krysinska et al., 2007). The experiments described above need to be reproduced to make sure that what we observed is true. These preliminary findings however indicate that MLL2 may be required to establish a correct transcriptional program during haemopoietic differentiation by either inducing a primed chromatin state in ES cells and/or by maintaining transcription of specific genes that drive the cells toward the haemopoietic lineages. If this holds true, it also suggests that while there is functional redundancy at some developmental stages, this redundancy may be lost later in development where correct expression of a target gene would solely depend on a specific H3K4 methyltransferase.

4.3.3. MLL2 in haemopoiesis – future directions

Our results suggest that MLL2 is required for the correct establishment of the haemopoietic system. However to firmly establish a clear role for MLL2 in this process further experimentation is required. Initially we need to repeat out blast culture expression analyses over a longer time-course. The panel of genes included needs to be expanded to include other key haemopoietic and possibly endothelial regulators.

To circumvent problems with decreased numbers of particular cell populations we propose to repeat the same assays using FACS purified cells, based on their C-KIT, CD41, TIE2 profile. Expression analyses in purified cells will provide further insight into the role of MLL2 in establishing the haemopoietic transcriptional program. Ideally, these populations will be examined using mRNA expression microarrays or RNA high throughput sequencing to obtain a general picture of MLL2 regulated transcription during haemopoietic development. It is absolutely necessary to repeat these experiments with *Mll2*$^{F/+}$ cells to exclude any OHT and/or CreERT2 induced effects.

Additionally, we would like to couple these analyses with genome-wide H3K4me$_3$, H3K27me$_3$ chromatin immunoprecipitation and DNA methylation studies to investigate whether deregulation of haemopoiesis-associated genes follows our proposed model for *Magoh2* regulation.

5. REFERENCES

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. Cell *103*, 667-678.

Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J.C., Suzuki, H., Daub, C.O., Hayashizaki, Y., and Lenhard, B. (2009). Transcriptional features of genomic regulatory blocks. Genome Biol *10*, R38.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proceedings of the National Academy of Sciences of the United States of America *51*, 786-794.

Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell *42*, 599-610.

Almer, A., Rudolph, H., Hinnen, A., and Horz, W. (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. The EMBO journal *5*, 2689-2696.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS biology *8*.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B.*, et al.* (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.

Ansari, K.I., Mishra, B.P., and Mandal, S.S. (2008). Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. Biochimica et biophysica acta *1779*, 66-73.

Apostolou, E., and Thanos, D. (2008). Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. Cell *134*, 85-96.

Archer, T.K., Cordingley, M.G., Wolford, R.G., and Hager, G.L. (1991). Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. Mol Cell Biol *11*, 688-698.

Austenaa, L., Barozzi, I., Chronowska, A., Termanini, A., Ostuni, R., Prosperini, E., Stewart, A.F., Testa, G., and Natoli, G. (2012). The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. Immunity *36*, 572-585.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1979). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med *149*, 297-326.

Axel, R. (1975). Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. Biochemistry *14*, 2921-2925.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol Cell Biol *24*, 10470-10478.

Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M.P., and Slany, R.K. (2009). Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815-823.

Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. Genes Dev *16*, 3186-3198.

Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., and Mills, A.A. (2007). CHD5 is a tumor suppressor at human 1p36. Cell *128*, 459-475.

Bainton, D.F. (1981). The discovery of lysosomes. The Journal of cell biology *91*, 66s-76s.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381-395.

Baxter, J.D., Rousseau, G.G., Benson, M.C., Garcea, R.L., Ito, J., and Tomkins, G.M. (1972). Role of DNA and specific cytoplasmic receptors in glucocorticoid action. Proceedings of the National Academy of Sciences of the United States of America *69*, 1892-1896.

Belikov, S., Holmqvist, P.H., Astrand, C., and Wrange, O. (2004). Nuclear factor 1 and octamer transcription factor 1 binding preset the chromatin structure of the mouse mammary tumor virus promoter for hormone induction. The Journal of biological chemistry *279*, 49857-49867.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984). Two regulatory elements for immunoglobulin kappa light chain gene expression. Proceedings of the National Academy of Sciences of the United States of America *81*, 7041-7045.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. The EMBO journal *7*, 1749-1756.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nature structural & molecular biology *14*, 1008-1016.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic acids research *8*, 1499-1504.

Bird, A.P., Taggart, M.H., Nicholls, R.D., and Higgs, D.R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. The EMBO journal *6*, 999-1004.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic acids research *30*, 958-965.

Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. Molecular cell *36*, 970-983.

Bohmann, D., Bos, T.J., Admon, A., Nishimura, T., Vogt, P.K., and Tjian, R. (1987). Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science *238*, 1386-1392.

Bonifer, C., and Bowen, D.T. (2010). Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine. Expert reviews in molecular medicine *12*, e6.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Boumber, Y.A., Kondo, Y., Chen, X., Shen, L., Guo, Y., Tellez, C., Estecio, M.R., Ahmed, S., and Issa, J.P. (2008). An Sp1/Sp3 binding polymorphism confers methylation protection. PLoS Genet *4*, e1000162.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brady, J., and Khoury, G. (1985). trans Activation of the simian virus 40 late transcription unit by T-antigen. Mol Cell Biol *5*, 1391-1399.

Bram, R.J., and Kornberg, R.D. (1985). Specific protein binding to far upstream activating sequences in polymerase II promoters. Proceedings of the National Academy of Sciences of the United States of America *82*, 43-47.

Brandeis, M., Ariel, M., and Cedar, H. (1993). Dynamics of DNA methylation during development. BioEssays : news and reviews in molecular, cellular and developmental biology *15*, 709-713.

Braunschweig, U., Hogan, G.J., Pagie, L., and van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. The EMBO journal *28*, 3635-3645.

Bres, V., Yoh, S.M., and Jones, K.A. (2008). The multi-tasking P-TEFb complex. Current opinion in cell biology *20*, 334-340.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev *15*, 3286-3295.

Brivanlou, A.H., and Darnell, J.E., Jr. (2002). Signal transduction and the control of gene expression. Science *295*, 813-818.

Brookes, P., and Lawley, P.D. (1963). Effects of Alkylating Agents on T2 and T4 Bacteriophages. The Biochemical journal *89*, 138-144.

Brownell, J.E., and Allis, C.D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proceedings of the National Academy of Sciences of the United States of America *92*, 6364-6368.

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nature immunology *4*, 235-240.

Bungard, D., Fuerth, B.J., Zeng, P.Y., Faubert, B., Maas, N.L., Viollet, B., Carling, D., Thompson, C.B., Jones, R.G., and Berger, S.L. (2010). Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. Science *329*, 1201-1205.

Buratowski, S. (1994). The basics of basal transcription by RNA polymerase II. Cell *77*, 1-3.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev *10*, 711-724.

Cairns, B.R. (2005). Chromatin remodeling complexes: strength in diversity, precision through specialization. Current opinion in genetics & development *15*, 185-190.

Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. Nature structural & molecular biology *14*, 989-996.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Current opinion in genetics & development *14*, 155-164.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Carrera, I., and Treisman, J.E. (2008). Message in a nucleus: signaling to the transcriptional machinery. Current opinion in genetics & development *18*, 397-403.

Chan, H.M., and La Thangue, N.B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. Journal of cell science *114*, 2363-2373.

Chang, P.Y., Hom, R.A., Musselman, C.A., Zhu, L., Kuo, A., Gozani, O., Kutateladze, T.G., and Cleary, M.L. (2010). Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. Journal of molecular biology *400*, 137-144.

Chen, T., and Li, E. (2004). Structure and function of eukaryotic DNA methyltransferases. Current topics in developmental biology *60*, 55-89.

Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T.G., Simpson, M., Mao, Q., Pan, C.H., Dai, S.*, et al.* (2006). Structural insights into histone demethylation by JMJD2 family members. Cell *125*, 691-702.

Cheng, T., Shen, H., Giokas, D., Gere, J., Tenen, D.G., and Scadden, D.T. (1996). Temporal mapping of gene expression levels during the differentiation of individual primary hematopoietic cells. Proceedings of the National Academy of Sciences of the United States of America *93*, 13158-13163.

Chow, A., Brown, B.D., and Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. Nature reviews Immunology *11*, 788-798.

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO reports *6*, 354-360.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annual review of biochemistry *78*, 273-304.

Clark, R.J., and Felsenfeld, G. (1971). Structure of chromatin. Nature: New biology *229*, 101-106.

Clayton, A.L., Rose, S., Barratt, M.J., and Mahadevan, L.C. (2000). Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. The EMBO journal *19*, 3714-3726.

Clever, U., and Ellgaard, E.G. (1970). Puffing and histone acetylation in polytene chromosomes. Science *169*, 373-374.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev *26*, 1714-1728.

Cockerill, P.N. (2000). Identification of DNaseI hypersensitive sites within nuclei. Methods Mol Biol *130*, 29-46.

Cockerill, P.N. (2011). Structure and function of active chromatin and DNase I hypersensitive sites. The FEBS journal *278*, 2182-2210.

Collins, E.C., and Rabbitts, T.H. (2002). The promiscuous MLL gene links chromosomal translocations to cellular differentiation and tumour tropism. Trends Mol Med *8*, 436-442.

Cordingley, M.G., and Hager, G.L. (1988). Binding of multiple factors to the MMTV promoter in crude and fractionated nuclear extracts. Nucleic acids research *16*, 609-628.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Corona, D.F., and Tamkun, J.W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochimica et biophysica acta *1677*, 113-119.

Cosgrove, M.S., and Patel, A. (2010). Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. The FEBS journal *277*, 1832-1842.

Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature *274*, 775-780.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J*., et al.* (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545-553.

Damm, F., Nguyen-Khac, F., Fontenay, M., and Bernard, O.A. (2012). Spliceosome and other novel mutations in chronic lymphocytic leukemia, and myeloid malignancies. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK.

Daser, A., and Rabbitts, T.H. (2005). The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. Semin Cancer Biol *15*, 175-188.

de Duve, C. (1996). The peroxisome in retrospect. Annals of the New York Academy of Sciences *804*, 1-10.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Deaton, A.M., Webb, S., Kerr, A.R., Illingworth, R.S., Guy, J., Andrews, R., and Bird, A. (2011). Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res *21*, 1074-1086.

DeKoter, R.P., Walsh, J.C., and Singh, H. (1998). PU.1 regulates both cytokine-dependent proliferation and differentiation of granulocyte/macrophage progenitors. The EMBO journal *17*, 4456-4468.

Demers, C., Chaturvedi, C.P., Ranish, J.A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F.J., Groudine, M., and Brand, M. (2007). Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. Molecular cell *27*, 573-584.

Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev *19*, 2418-2423.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. Cell *32*, 695-706.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D., and Evans, G.A. (1992). A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. Nat Genet *2*, 113-118.

Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. The Journal of biological chemistry *280*, 17986-17991.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of embryology and experimental morphology *87*, 27-45.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nature structural & molecular biology *13*, 713-719.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873-885.

Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M., and Coulombe, B. (2000). Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. Mol Cell Biol *20*, 8168-8177.

Driever, W., and Nusslein-Volhard, C. (1988). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell *54*, 95-104.

Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79-87.

Eberharter, A., and Becker, P.B. (2004). ATP-dependent nucleosome remodelling: factors and functions. Journal of cell science *117*, 3707-3711.

Edlund, T., Walker, M.D., Barr, P.J., and Rutter, W.J. (1985). Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5' flanking elements. Science *230*, 912-916.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic acids research *10*, 2709-2721.

Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proceedings of the National Academy of Sciences of the United States of America *105*, 19732-19737.

Engstrom, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., and Lenhard, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res *17*, 1898-1908.

Ephrussi, A., Church, G.M., Tonegawa, S., and Gilbert, W. (1985). B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science *227*, 134-140.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H.*, et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology *15*, 1176-1183.

Ernst, P., Fisher, J.K., Avery, W., Wade, S., Foy, D., and Korsmeyer, S.J. (2004). Definitive hematopoiesis requires the mixed-lineage leukemia gene. Developmental cell *6*, 437-443.

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol Cell Biol *21*, 2249-2258.

Eskiw, C.H., and Fraser, P. (2011). Ultrastructural study of transcription factories in mouse erythroblasts. Journal of cell science *124*, 3676-3683.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Evans, R., Fairley, J.A., and Roberts, S.G. (2001). Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev *15*, 2945-2949.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet *4*, e1000116.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development *130*, 4217-4227.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188-194.

Felsenfeld, G., and McGhee, J.D. (1986). Structure of the 30 nm chromatin fiber. Cell *44*, 375-377.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I.*, et al.* (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res *22*, 2399-2408.

FitzGerald, K.T., and Diaz, M.O. (1999). MLL2: A new mammalian member of the trx/MLL family of genes. Genomics *59*, 187-192.

Fletcher, T.M., Xiao, N., Mautino, G., Baumann, C.T., Wolford, R., Warren, B.S., and Hager, G.L. (2002). ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. Mol Cell Biol *22*, 3255-3263.

Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev *15*, 1783-1795.

Fraser, N.W., Sehgal, P.B., and Darnell, J.E. (1978). DRB-induced premature termination of late adenovirus transcription. Nature *272*, 590-593.

Fryer, C.J., and Archer, T.K. (1998). Chromatin remodelling by the glucocorticoid receptor requires the BRG1 complex. Nature *393*, 88-91.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003a). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic acids research *31*, 2305-2312.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003b). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. The Journal of biological chemistry *278*, 4035-4040.

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic acids research *5*, 3157-3170.

Gall, J.G. (1981). Chromosome structure and the C-value paradox. The Journal of cell biology *91*, 3s-14s.

Gallwitz, D., and Sekeris, C.E. (1969). The acetylation of histones of rat liver nuclei in vitro by acetyl-CoA. Hoppe-Seyler's Zeitschrift fur physiologische Chemie *350*, 150-154.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell *33*, 717-728.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984-3989.

Glaser, S., Lubitz, S., Loveland, K.L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K.*, et al.* (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics Chromatin *2*, 5.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423-1432.

Goo, Y.H., Sohn, Y.C., Kim, D.H., Kim, S.W., Kang, M.J., Jung, D.J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T.*, et al.* (2003). Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Mol Cell Biol *23*, 140-149.

Goodrich, J.A., and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell *77*, 145-156.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet *11*, 549-558.

Grange, T., Bertrand, E., Espinas, M.L., Fromont-Racine, M., Rigaud, G., Roux, J., and Pictet, R. (1997). In vivo footprinting of the interaction of proteins with DNA and RNA. Methods *11*, 151-163.

Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biological reviews of the Cambridge Philosophical Society *76*, 65-101.

Grigoryev, S.A., and Woodcock, C.L. (2012). Chromatin organization - The 30nm fiber. Experimental cell research *318*, 1448-1455.

Gupta, A., Guerin-Peyrou, T.G., Sharma, G.G., Park, C., Agarwal, M., Ganju, R.K., Pandita, S., Choi, K., Sukumar, S., Pandita, R.K., *et al.* (2008). The mammalian ortholog of Drosophila MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. Mol Cell Biol *28*, 397-409.

Gurley, L.R., Walters, R.A., and Tobey, R.A. (1974). Cell cycle-specific changes in histone phosphorylation associated with cell proliferation and chromosome condensation. The Journal of cell biology *60*, 356-364.

Hager, L.J., and Palmiter, R.D. (1981). Transcriptional regulation of mouse liver metallothionein-I gene by glucocorticoids. Nature *291*, 340-342.

Hall, B.D., Clarkson, S.G., and Tocchini-Valentini, G. (1982). Transcription initiation of eucaryotic transfer RNA genes. Cell *29*, 3-5.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. Cell *138*, 129-145.

Harrington, M.A., Jones, P.A., Imagawa, M., and Karin, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proceedings of the National Academy of Sciences of the United States of America *85*, 2066-2070.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic acids research *40*, 4841-4849.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Hedley, M.L., Amrein, H., and Maniatis, T. (1995). An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. Proceedings of the National Academy of Sciences of the United States of America *92*, 11524-11528.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Henikoff, S. (2009). Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. Nat Genet *41*, 865-866.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends in genetics : TIG *27*, 389-396.

Hess, J.L., Yu, B.D., Li, B., Hanson, R., and Korsmeyer, S.J. (1997). Defects in yolk sac hematopoiesis in Mll-null embryos. Blood *90*, 1799-1806.

Heuchel, R., Matthias, P., and Schaffner, W. (1989). Two closely spaced promoters are equally activated by a remote enhancer: evidence against a scanning model for enhancer action. Nucleic acids research *17*, 8931-8947.

Himes, S.R., Tagoh, H., Goonetilleke, N., Sasmono, T., Oceandy, D., Clark, R., Bonifer, C., and Hume, D.A. (2001). A highly conserved c-fms gene intronic element controls macrophage-specific and regulated expression. Journal of leukocyte biology *70*, 812-820.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. J Biochem *141*, 601-608.

Hitchins, M.P., Rapkins, R.W., Kwok, C.T., Srivastava, S., Wong, J.J., Khachigian, L.M., Polly, P., Goldblatt, J., and Ward, R.L. (2011). Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. Cancer cell *20*, 200-213.

Hofemeister, H., Ciotta, G., Fu, J., Seibert, P.M., Schulz, A., Maresca, M., Sarov, M., Anastassiadis, K., and Stewart, A.F. (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods *53*, 437-452.

Holler, M., Westin, G., Jiricny, J., and Schaffner, W. (1988). Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev *2*, 1127-1135.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., *et al.* (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. Blood *114*, 299-309.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H.*, et al.* (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res *21*, 182-192.

Hsieh, J.J., Cheng, E.H., and Korsmeyer, S.J. (2003). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. Cell *115*, 293-303.

Hughes, C.M., Rozenblatt-Rosen, O., Milne, T.A., Copeland, T.D., Levine, S.S., Lee, J.C., Hayes, D.N., Shanmugam, K.S., Bhattacharjee, A., Biondi, C.A.*, et al.* (2004). Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Molecular cell *13*, 587-597.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J.*, et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS biology *6*, e22.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet *6*.

Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. Nature *306*, 591-593.

Iniguez-Lluhi, J.A. (2006). For a healthy histone code, a little SUMO in the tail keeps the acetyl away. ACS chemical biology *1*, 204-206.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jin, S.G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic acids research *38*, e125.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-191.

Jonsson, Z.O., Jha, S., Wohlschlegel, J.A., and Dutta, A. (2004). Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Molecular cell *16*, 465-477.

Joyner, A.L., Skarnes, W.C., and Rossant, J. (1989). Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells. Nature *338*, 153-156.

Jude, C.D., Climer, L., Xu, D., Artinger, E., Fisher, J.K., and Ernst, P. (2007). Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. Cell stem cell *1*, 324-337.

Jurgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature *316*, 153-155.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. Current opinion in cell biology *20*, 253-259.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol *339*, 225-229.

Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev *24*, 2615-2620.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Kanhere, A., Viiri, K., Araujo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., *et al.* (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Molecular cell *38*, 675-688.

Karpf, A.R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer research *65*, 8635-8639.

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., and Cairns, B.R. (2004). Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. The EMBO journal *23*, 1348-1359.

Kataoka, N., Diem, M.D., Kim, V.N., Yong, J., and Dreyfuss, G. (2001). Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. The EMBO journal *20*, 6424-6433.

Kazazian, H.H., Jr., and Moran, J.V. (1998). The impact of L1 retrotransposons on the human genome. Nat Genet *19*, 19-24.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480-485.

Kim, J., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459-471.

Kinoshita, T., Fujita, M., and Maeda, Y. (2008). Biosynthesis, remodelling and functions of mammalian GPI-anchored proteins: recent progress. J Biochem *144*, 287-294.

Kleff, S., Andrulis, E.D., Anderson, C.W., and Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. The Journal of biological chemistry *270*, 24674-24677.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS biology *2*, E131.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions-long distance calls. Current opinion in genetics & development *22*, 79-85.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., *et al.* (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Molecular cell *11*, 721-729.

Krysinska, H., Hoogenkamp, M., Ingram, R., Wilson, N., Tagoh, H., Laslo, P., Singh, H., and Bonifer, C. (2007). A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c-fms gene. Mol Cell Biol *27*, 878-887.

Kuff, E.L., and Lueders, K.K. (1988). The intracisternal A-particle gene family: structure and functional aspects. Advances in cancer research *51*, 183-276.

Kutach, A.K., and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol *20*, 4754-4764.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev *12*, 34-44.

Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547-553.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature *457*, 892-895.

Lee, J., Kim, D.H., Lee, S., Yang, Q.H., Lee, D.K., Lee, S.K., Roeder, R.G., and Lee, J.W. (2009). A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proceedings of the National Academy of Sciences of the United States of America *106*, 8513-8518.

Lee, J., Saha, P.K., Yang, Q.H., Lee, S., Park, J.Y., Suh, Y., Lee, S.K., Chan, L., Roeder, R.G., and Lee, J.W. (2008). Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. Proceedings of the National Academy of Sciences of the United States of America *105*, 19229-19234.

Lee, J.H., and Skalnik, D.G. (2002). CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. The Journal of biological chemistry *277*, 42259-42267.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. The Journal of biological chemistry *280*, 41725-41731.

Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. The Journal of biological chemistry *276*, 44669-44676.

Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.

Lee, S., Lee, D.K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.Y., Lee, S.K., Roeder, R.G., and Lee, J.W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proceedings of the National Academy of Sciences of the United States of America *103*, 15392-15397.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics *34*, 77-137.

Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987a). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature *325*, 368-372.

Lee, W., Mitchell, P., and Tjian, R. (1987b). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell *49*, 741-752.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal *23*, 4051-4060.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13*, 233-245.

Levinger, L., and Varshavsky, A. (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. Cell *28*, 375-385.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Lewis, E.B. (1982). Control of body segment differentiation in Drosophila by the bithorax gene complex. Progress in clinical and biological research *85 Pt A*, 269-288.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. The Journal of biological chemistry *278*, 8897-8903.

Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. The Journal of biological chemistry *277*, 49383-49388.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev *18*, 1606-1617.

Lomvardas, S., and Thanos, D. (2001). Nucleosome sliding via TBP DNA binding in vivo. Cell *106*, 685-696.

Long, J., Zuo, D., and Park, M. (2005). Pc2-mediated sumoylation of Smad-interacting protein 1 attenuates transcriptional repression of E-cadherin. The Journal of biological chemistry *280*, 35477-35489.

Lubitz, S., Glaser, S., Schaft, J., Stewart, A.F., and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. Mol Biol Cell *18*, 2356-2366.

Lucey, M.J., Chen, D., Lopez-Garcia, J., Hart, S.M., Phoenix, F., Al-Jehani, R., Alao, J.P., White, R., Kindle, K.B., Losson, R.*, et al.* (2005). T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. Nucleic acids research *33*, 6393-6404.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. The Journal of biological chemistry *286*, 35334-35338.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761-772.

Mansergh, F.C., Daly, C.S., Hurley, A.L., Wride, M.A., Hunter, S.M., and Evans, M.J. (2009). Gene expression profiles during early differentiation of mouse embryonic stem cells. BMC developmental biology *9*, 5.

Margaritis, T., and Holstege, F.C. (2008). Poised RNA polymerase II gives pause for thought. Cell *133*, 581-584.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. The Journal of biological chemistry *271*, 27176-27183.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol *12*, 2078-2090.

Martens, J.A., and Winston, F. (2003). Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Current opinion in genetics & development *13*, 136-142.

Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO reports *5*, 490-496.

Mathis, D.J., and Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature *290*, 310-315.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America *74*, 560-564.

McBride, H.M., Neuspiel, M., and Wasiak, S. (2006). Mitochondria: more than just a powerhouse. Current biology : CB *16*, R551-560.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D., and Felsenfeld, G. (1981). A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell *27*, 45-55.

McGinty, R.K., Kim, J., Chatterjee, C., Roeder, R.G., and Muir, T.W. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature *453*, 812-816.

McMahon, K.A., Hiew, S.Y., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A.J., Kioussis, D., Williams, O., and Brady, H.J. (2007). Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. Cell stem cell *1*, 338-345.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Molecular cell *1*, 277-287.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell *10*, 105-116.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G.*, et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Molecular cell *10*, 1107-1117.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005a). MLL associates specifically with a subset of transcriptionally active target genes. Proceedings of the National Academy of Sciences of the United States of America *102*, 14765-14770.

Milne, T.A., Hughes, C.M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepp, R.W., Krankel, C., Livolsi, V.A., Gibbs, D.*, et al.* (2005b). Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. Proceedings of the National Academy of Sciences of the United States of America *102*, 749-754.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G.*, et al.* (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Molecular cell *38*, 853-863.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090-1097.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.

Mo, R., Rao, S.M., and Zhu, Y.J. (2006). Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. The Journal of biological chemistry *281*, 15714-15720.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic acids research *9*, 6047-6068.

Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N., and Mellor, J. (2003). Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. Cell *115*, 425-435.

Morrison, A.J., Highland, J., Krogan, N.J., Arbel-Eden, A., Greenblatt, J.F., Haber, J.E., and Shen, X. (2004). INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. Cell *119*, 767-775.

Mueller-Storm, H.P., Sogo, J.M., and Schaffner, W. (1989). An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. Cell *58*, 767-777.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D., and Cleary, M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proceedings of the National Academy of Sciences of the United States of America *99*, 90-94.

Nakabeppu, Y., Ryder, K., and Nathans, D. (1988). DNA binding activities of three murine Jun proteins: stimulation by Fos. Cell *55*, 907-915.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905-916.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular cell *11*, 709-719.

Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C.*, et al.* (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet *42*, 790-793.

Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kedinger, C., and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. Nucleic acids research *24*, 2924-2929.

Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: a structural view. Proceedings of the National Academy of Sciences of the United States of America *94*, 15-22.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong, A.S.J.*, et al.* (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188-4197.

Ohler, U., Liao, G.C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the Drosophila genome. Genome Biol *3*, RESEARCH0087.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Oliva, R., Bazett-Jones, D.P., Locklear, L., and Dixon, G.H. (1990). Histone hyperacetylation can induce unfolding of the nucleosome core particle. Nucleic acids research *18*, 2739-2747.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W.*, et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet *36*, 1065-1071.

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. The EMBO journal *19*, 6141-6149.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nature biotechnology *25*, 244-248.

Palade, G. (1975). Intracellular aspects of the process of protein synthesis. Science *189*, 867.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Parthun, M.R. (2007). Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319-5328.

Patel, S.R., Kim, D., Levitan, I., and Dressler, G.R. (2007). The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. Developmental cell *13*, 580-592.

Pearson, S., Sroczynska, P., Lacaud, G., and Kouskoff, V. (2008). The stepwise specification of embryonic stem cells to hematopoietic fate is driven by sequential exposure to Bmp4, activin A, bFGF and VEGF. Development *135*, 1525-1535.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Molecular cell *23*, 297-305.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Current biology : CB *14*, R546-551.

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Stoerker, J., Genuardi, M., Yeung, A.T., Matsumoto, Y., and Bellacosa, A. (2000). Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. The Journal of biological chemistry *275*, 32422-32429.

Pfeifer, G.P., and Riggs, A.D. (1991). Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev *5*, 1102-1113.

Phair, R.D., Gorski, S.A., and Misteli, T. (2004a). Measurement of dynamic protein binding to chromatin in vivo, using photobleaching microscopy. Methods in enzymology *375*, 393-414.

Phair, R.D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M., and Misteli, T. (2004b). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. Mol Cell Biol *24*, 6393-6402.

Pogo, B.G., Allfrey, V.G., and Mirsky, A.E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. Proceedings of the National Academy of Sciences of the United States of America *55*, 805-812.

Price, D.H. (2008). Poised polymerases: on your mark...get set...go! Molecular cell *30*, 7-10.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev *15*, 1613-1618.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Pullirsch, D., Hartel, R., Kishimoto, H., Leeb, M., Steiner, G., and Wutz, A. (2010). The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development *137*, 935-943.

Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS Genet *7*, e1001274.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reeves, R., and Nissen, M.S. (1990). The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. The Journal of biological chemistry *265*, 8573-8582.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Robinson, P.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. Journal of molecular biology *381*, 816-825.

Robinson, P.J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Current opinion in structural biology *16*, 336-343.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. The EMBO journal *24*, 2354-2366.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. The EMBO journal *20*, 7137-7148.

Rozenblatt-Rosen, O., Hughes, C.M., Nannepaga, S.J., Shanmugam, K.S., Copeland, T.D., Guszczynski, T., Resau, J.H., and Meyerson, M. (2005). The parafibromin tumor suppressor protein is part of a human Paf1 complex. Mol Cell Biol *25*, 612-620.

Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C.M., Mazo, A., and Canaani, E. (1998). The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. Proceedings of the National Academy of Sciences of the United States of America *95*, 4152-4157.

Ruault, M., Brun, M.E., Ventura, M., Roizes, G., and De Sario, A. (2002). MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. Gene *284*, 73-81.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell *25*, 15-30.

Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M., and Felsenfeld, G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. The EMBO journal *19*, 2315-2322.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nature reviews Molecular cell biology *7*, 557-567.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America *103*, 1412-1417.

Scharf, A.N., Barth, T.K., and Imhof, A. (2009). Establishment of histone modifications after chromatin assembly. Nucleic acids research *37*, 5032-5040.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics *90*, 314-323.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nature reviews Molecular cell biology *12*, 799-814.

Schwabish, M.A., and Struhl, K. (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. Mol Cell Biol *24*, 10111-10117.

Schwartz, B.E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes Dev *19*, 804-814.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9-22.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends in genetics : TIG *25*, 335-343.

Shalaby, F., Ho, J., Stanford, W.L., Fischer, K.D., Schuh, A.C., Schwartz, L., Bernstein, A., and Rossant, J. (1997). A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis. Cell *89*, 981-990.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shinkai, Y. (2007). Regulation and function of H3K9 methylation. Sub-cellular biochemistry *41*, 337-350.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T.*, et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proceedings of the National Academy of Sciences of the United States of America *100*, 15776-15781.

Simpson, R.T. (1978). Structure of chromatin containing extensively acetylated H3 and H4. Cell *13*, 691-699.

Singer, J., Roberts-Ems, J., and Riggs, A.D. (1979). Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II. Science *203*, 1019-1021.

Singer, S.J., and Nicolson, G.L. (1972). The fluid mosaic model of the structure of cell membranes. Science *175*, 720-731.

Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T.*, et al.* (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337-342.

Slany, R.K. (2005). When epigenetics kills: MLL fusion proteins in leukemia. Hematological oncology *23*, 1-9.

Slany, R.K. (2009). The molecular biology of mixed lineage leukemia. Haematologica *94*, 984-993.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103-113.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual review of biochemistry *72*, 449-479.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes Dev *25*, 661-672.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230-234.

Soloaga, A., Thomson, S., Wiggin, G.R., Rampersaud, N., Dyson, M.H., Hazzalin, C.A., Mahadevan, L.C., and Arthur, J.S. (2003). MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. The EMBO journal *22*, 2788-2797.

Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A., and Baltimore, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature *323*, 640-643.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 3418-3422.

Stewart, C.L., Stuhlmann, H., Jahner, D., and Jaenisch, R. (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 4098-4102.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev *5*, 820-826.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Tagoh, H., Cockerill, P.N., and Bonifer, C. (2006). In vivo genomic footprinting using LM-PCR methods. Methods Mol Biol *325*, 285-314.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takeda, S., Chen, D.Y., Westergard, T.D., Fisher, J.K., Rubens, J.A., Sasagawa, S., Kan, J.T., Korsmeyer, S.J., Cheng, E.H., and Hsieh, J.J. (2006). Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. Genes Dev *20*, 2397-2409.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nature reviews Molecular cell biology *11*, 264-275.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell *68*, 561-572.

Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proceedings of the National Academy of Sciences of the United States of America *103*, 6629-6634.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Thanos, D., Mavrothalassitis, G., and Papamatheakis, J. (1988). Multiple regulatory regions on the 5' side of the mouse E alpha gene. Proceedings of the National Academy of Sciences of the United States of America *85*, 3075-3079.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thomas, K.R., Deng, C., and Capecchi, M.R. (1992). High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. Mol Cell Biol *12*, 2919-2923.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D.*, et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Tkachuk, D.C., Kohler, S., and Cleary, M.L. (1992). Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691-700.

Treisman, R., and Maniatis, T. (1985). Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. Nature *315*, 73-75.

Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. Current opinion in genetics & development *5*, 190-196.

Trifonov, E.N., and Sussman, J.L. (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. Proceedings of the National Academy of Sciences of the United States of America *77*, 3816-3820.

Truss, M., Bartsch, J., Schelbert, A., Hache, R.J., and Beato, M. (1995). Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter in vivo. The EMBO journal *14*, 1737-1751.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tsukiyama, T., Becker, P.B., and Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature *367*, 525-532.

Turker, M.S. (2002). Gene silencing in mammalian cells and the spread of DNA methylation. Oncogene *21*, 5388-5393.

Turner, B.M. (2012). The adjustable nucleosome: an epigenetic signaling module. Trends in genetics : TIG.

Tyagi, S., Chabes, A.L., Wysocka, J., and Herr, W. (2007). E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. Molecular cell *27*, 107-119.

Tyagi, S., and Herr, W. (2009). E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases. The EMBO journal *28*, 3185-3195.

Vakoc, C.R., Wen, Y.Y., Gibbs, R.A., Johnstone, C.N., Rustgi, A.K., and Blobel, G.A. (2009). Low frequency of MLL3 mutations in colorectal carcinoma. Cancer genetics and cytogenetics *189*, 140-141.

van Attikum, H., Fritsch, O., Hohn, B., and Gasser, S.M. (2004). Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. Cell *119*, 777-788.

van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure *16*, 1245-1256.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. Proceedings of the National Academy of Sciences of the United States of America *79*, 1073-1077.

Vaute, O., Nicolas, E., Vandel, L., and Trouche, D. (2002). Functional and physical interaction between the histone methyl transferase Suv39H1 and histone deacetylases. Nucleic acids research *30*, 475-481.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes Dev *25*, 1583-1588.

Vidali, G., Boffa, L.C., Bradbury, E.M., and Allfrey, V.G. (1978). Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. Proceedings of the National Academy of Sciences of the United States of America *75*, 2239-2243.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Walter, W., Clynes, D., Tang, Y., Marmorstein, R., Mellor, J., and Berger, S.L. (2008). 14-3-3 interaction with histone H3 involves a dual modification pattern of phosphoacetylation. Mol Cell Biol *28*, 2840-2849.

Walters, M.C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., and Martin, D.I. (1995). Enhancers increase the probability but not the level of gene expression. Proceedings of the National Academy of Sciences of the United States of America *92*, 7125-7129.

Walters, M.C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M., and Martin, D.I. (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. Genes Dev *10*, 185-195.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M.*, et al.* (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol Cell Biol *29*, 6074-6085.

Wang, X.X., Fu, L., Li, X., Wu, X., Zhu, Z., and Dong, J.T. (2011). Somatic mutations of the mixed-lineage leukemia 3 (MLL3) gene in primary breast cancers. Pathology oncology research : POR *17*, 429-433.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y.*, et al.* (2004b). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279-283.

Wasylyk, B., and Chambon, P. (1983). Potentiator effect of the SV40 72-bp repeat on initiation of transcription from heterologous promoter elements. Cold Spring Harbor symposia on quantitative biology *47 Pt 2*, 921-934.

Watanabe, R., Murakami, Y., Marmor, M.D., Inoue, N., Maeda, Y., Hino, J., Kangawa, K., Julius, M., and Kinoshita, T. (2000). Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. The EMBO journal *19*, 4402-4411.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev *2*, 1136-1143.

Weake, V.M., and Workman, J.L. (2008). Histone ubiquitination: triggering gene activity. Molecular cell *29*, 653-663.

Weber, F., and Schaffner, W. (1985). Simian virus 40 enhancer increases RNA polymerase density within the linked gene. Nature *315*, 75-77.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Wedeen, C., Harding, K., and Levine, M. (1986). Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in Drosophila. Cell *44*, 739-748.

Weil, P.A., Luse, D.S., Segall, J., and Roeder, R.G. (1979). Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell *18*, 469-484.

Weiss, E., Ruhlmann, C., and Oudet, P. (1986). Transcriptionally active SV40 minichromosomes are restriction enzyme sensitive and contain a nucleosome-free origin region. Nucleic acids research *14*, 2045-2058.

Wilhelm, J.A., and McCarty, K.S. (1970). Partial characterization of the histones and histone acetylation in cell cultures. Cancer research *30*, 409-417.

Wobus, A.M., Grosse, R., and Schoneich, J. (1988). Specific effects of nerve growth factor on the differentiation pattern of mouse embryonic stem cells in vitro. Biomedica biochimica acta *47*, 965-973.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annual review of biochemistry *67*, 545-579.

Workman, J.L., and Roeder, R.G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell *51*, 613-622.

Wotton, D., and Merrill, J.C. (2007). Pc2 and SUMOylation. Biochemical Society transactions *35*, 1401-1404.

Wright, D.E., Wang, C.Y., and Kao, C.F. (2012). Histone ubiquitylation and chromatin dynamics. Frontiers in bioscience : a journal and virtual library *17*, 1051-1078.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nature reviews Molecular cell biology *11*, 607-620.

Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N., and Herr, W. (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes Dev *17*, 896-911.

Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat Genet *26*, 229-232.

Yagi, H., Deguchi, K., Aono, A., Tani, Y., Kishimoto, T., and Komori, T. (1998). Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. Blood *92*, 108-117.

Yagi, S., Hirabayashi, K., Sato, S., Li, W., Takahashi, Y., Hirakawa, T., Wu, G., Hattori, N., Ohgane, J., Tanaka, S.*, et al.* (2008). DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. Genome Res *18*, 1969-1978.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. (2005). Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. Gene *350*, 129-136.

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proceedings of the National Academy of Sciences of the United States of America *97*, 11303-11306.

Yang, Z., He, N., and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Mol Cell Biol *28*, 967-976.

Yao, T.P., Oh, S.P., Fuchs, M., Zhou, N.D., Ch'ng, L.E., Newsome, D., Bronson, R.T., Li, E., Livingston, D.M., and Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell *93*, 361-372.

Yie, J., Senger, K., and Thanos, D. (1999). Mechanism by which the IFN-beta enhanceosome activates transcription. Proceedings of the National Academy of Sciences of the United States of America *96*, 13108-13113.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in genetics : TIG *13*, 335-340.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Mol Cell Biol *24*, 5639-5649.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003a). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Molecular cell *12*, 723-734.

Yoon, J.H., Iwai, S., O'Connor, T.R., and Pfeifer, G.P. (2003b). Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. Nucleic acids research *31*, 5399-5404.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505-508.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512-1516.

Zernicka-Goetz, M. (2005). Cleavage pattern and emerging asymmetry of the mouse embryo. Nature reviews Molecular cell biology *6*, 919-928.

Zhang, L., and Gralla, J.D. (1989). Micrococcal nuclease as a probe for bound and distorted DNA in lac transcription and repression complexes. Nucleic acids research *17*, 5017-5028.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122-1136.

Zylber, E.A., and Penman, S. (1971). Products of RNA polymerases in HeLa cell nuclei. Proceedings of the National Academy of Sciences of the United States of America *68*, 2861-2865.